

British research scrutinized again

The House of Lords is to look again at the organization of British science. It could do worse than reprint its proposals on the same subject five years ago.

THE frequent complaint that the British parliament knows nothing of the conduct of science is a lie. Although the House of Commons is no better endowed with technical people than other elected legislatures, the House of Lords, for a quarter of a century dominated by people with only a life tenancy of a title, is a different kettle of fish. Five years ago, the Lords Select Committee on Science and Technology undertook an important study of how the British government runs research and concluded that affairs were badly managed. Now, with further experience bearing out that general conclusion, the same committee (with different people) is about to embark on essentially the same task, and is asking for opinions by 14 February (in writing) on the questions it has posed for itself. Here are shortened forms of the questions: the accompanying telegraphic answers are included for the convenience of readers too occupied with more urgent tasks (rewriting research grant applications, for example).

Q. *How far should public support for science and technology be an objective of national policy?* **A.** As far as a piece of string is long. All governments agree that there must be some public support; most know that an arbitrary degree of support will be justified by its consequences. The governments of France (see p. 165), Japan, West Germany and the United States, after five years of generous support for science and technology, are able to point to progress of the kind they promised; the British government, starting from the proposition that British science used to be a profligate business, can now claim as justification the fact that previously academic researchers have commonly become suppliants to industry for research funds.

Q. *Could the present division of responsibility between the research councils be improved?* **A.** What responsibility? The theory, at the time of the Rothschild report in 1971, was that the Science and Engineering Research Council (SERC, then significantly only SRC), is alone exclusively concerned with the welfare of academic research; the three other science research councils (the Economic and Social Research Council is in a dog-house all of its own) had specific remits for research in agriculture, medicine and the natural environment. They are goal-directed organizations, but their programmes range from basic research to some operational activity (geological survey, for example).

Much has changed in the past five years. Under financial pressure, the three operational research councils have been cutting back on their in-house research. (University people are also cheaper.) These councils appear also to have been genuinely moved by the condition into which British academic research has sunk, and are doing what they can to divert resources from their own establishments towards universities. When applied projects can usefully be carried out by academics, there is no complaint. One danger, probably not yet serious, is that there will be an awkward overlapping of criteria applied to basic academic research. Another, more immediate, is that the Agricultural and Food Research Council has probably now been weakened to the point at which it cannot do its job.

The Natural Environment Research Council (NERC) is in a different kind of trouble, never having been given a coherent job

to do. The result is that it has a ragbag of functions, ranging from the support of academic earth science to the running of the British Geological Survey. The case for NERC's separate continued existence has vanished. (Lazy readers wishing to make the same point without being identified with the repetitious source of this opinion should either skip this question or say that the answer is "self-evident".)

Q. *What organizational changes are needed?* **A.** The fewest compatible with the survival of British research. There have been so many organizational changes so recently that stability would be welcome. Changes of substance are now more necessary than changes of form. But there is a need that some disinterested cabinet minister other than Mrs Margaret Thatcher should be asked to keep a watching brief on science, which is what the House of Lords committee recommended five years ago. Since then, the lack of political interest in science has patently allowed a once flourishing enterprise to be damaged, perhaps beyond repair.

If the case for a minister to look after science was strong five years ago, the intervening experience has only reinforced it. A strictly organizational benefit would be that the Advisory Board for the Research Councils, which exists merely to recommend a division of the annual science budget, would have a more pointed focus for its work. Subsuming NERC in other agencies would, by comparison, be a minor blessing.

Discerning readers will at this point recognize this question to be a trap. For all the constitutional importance of the House of Lords, it is not on this occasion bent on defending the entitlement of researchers in British laboratories to decent (let alone internationally competitive) salaries and other terms of service, but the personal penury that goes with a career in basic research will have to be tackled at some stage. What may worry the House of Lords (and should be keeping British researchers awake at nights) is the whiff of scandal that attends the quasi-autonomous conduct by British scientists of their own affairs.

Scandal? Nobody suggests that people have been stuffing public money into their pockets. The scandals that matter, and that damage science as well as the public purse, are more often errors of omission than of commission. There is, for example, the case of SERC's attempt, over four agonizing years, to reduce its spending on the infrastructure of astronomy; the committee charged with finding a solution, with the previous chairman of SERC at its head, has now apparently repeated the customary recommendation that nothing can be done. It seems not to be appreciated that indecisiveness on this scale plays into the hands of those in the British government, the Prime Minister in particular, who complain that basic research is effete. Maybe the new chairman of SERC, Professor Bill Mitchell, will be able to grasp the nettle that his predecessor has shrunk from, but the general question of how the research councils are managed needs attention. Professor Ralph Slatyer's argument (see *Nature* 319, 3; 1986) in comparable Australian circumstances, that the chief executive should be the appointed servant of the policy-making council, should be compulsory reading for the House of Lords committee.

Q. *Are there other sources of funds than government for applied*



research? A. Yes, but the question is badly put. Government spending on applied research is already substantial, chiefly through the government's own laboratories, especially defence laboratories; more should be spent by means of contracts with industrial and research organizations. What the House of Lords appears to have in mind is the possibility that industrial money from applied research might be used to prop up the research enterprise as a whole. Many universities have increased their incomes substantially by industrial contracts; the cost, in the diversion of creative people's interests, is not yet known. The chance that research council institutes may be able to save themselves by similar means is, however, much smaller.

Q. *How is the Rothschild principle working?* A. Not well at all, partly because the government that in 1971 enthusiastically welcomed the doctrine that applied research should be carried out by research councils, acting as contractors for customers identified as government departments, never gave the doctrine a chance. The notion that 10 per cent of the funds transferred in this way should be for basic research was never tried. Government departments have not been provided with chief scientists capable of operating the system, and have been free to renege on their commitments when it suited their convenience, causing mayhem in the research laboratories concerned.

Q. *Can defence research help civil research?* A. Yes, but only if the system is reorganized. The Prime Minister, Mrs Margaret Thatcher, and her now departed Defence Secretary, have never adequately delivered their promise, two years ago, to make defence research more open to British civil industry. Secrecy is only part of the problem. Lack of interest is a more serious obstacle. This is yet another reason why there should be a minister not merely to coordinate the government's spending on research but to develop a coherent policy on the subject.

The benefits of such a role are wider than mere coordination, although the value of a means of getting important questions decided cannot be underestimated. SERC may have procrastinated over the organization of astronomy, keeping those working at its establishments on tenterhooks, but the government is worse. Nothing has been heard of the proposal that Britain should bargain a reduced contribution to, or withdraw from, the European high-energy physics organization (CERN) since the government assumed responsibility from the Advisory Board for the Research Councils. Does anybody care what the consequences are for British academics, graduate students and potential graduate students who are vitally interested parties? On a smaller scale, but one not less vital for those whose careers are affected, the Ministry of Agriculture, Fisheries and Food has been brooding since last summer on an intelligent proposal that the Plant Breeding Institute, which develops new cereal varieties (among other things) should be joined in some manner to the National Seed Development Organization (which makes a profit for the Treasury by exploiting the same varieties. It is simply maladministration that these questions are neglected.

The question of money is important. The British government's policy for six years has been to keep the cost of civil science static. The enforced premature closure of once productive enterprises may have helped to balance the Treasury's books, but has otherwise been a diseconomy. It can only be a matter of time before the government (or its successor) has to change tack, and to behave like a spendthrift to undo some of the damage that has been done.

There is also the matter of people's morale to be considered. There is something in the belief that what is wrong with the British research enterprise is not the lack of money, serious though that may be, but the way in which continuing uncertainty has sapped people's belief that the research they do will be a foundation for even better things. In basic research, the time has gone, in Britain, when there was reasonable hope that modest beginnings could, if successful, grow into bigger enterprises. That is the sense that needs to be restored, more quickly than the supply of mere money. □

Agenda for Geneva

The resumption this week of the superpower talks is a time to ask what they can achieve.

As the Reagan-Gorbachev summit recedes in time, so do its problems. This week, the reassembly of the bilateral arms control negotiations at Geneva should bring them back abruptly. The US Strategic Defense Initiative (SDI) is bound to be high on the agenda; the summit participants agreed only to disagree about it. The first objective should be to prevent this divisive issue from wrecking the hard work that must be done on the control of strategic missiles. The second should be to confine the bargaining within the framework of what is realistically possible. Then, this year may even yield some tangible agreements.

This time, optimism is in order. That the summit happened, and did not finish up as a shouting match, is one good sign. That there is to be a second meeting later this year is another; the temptation to sign an agreement with some substance will be irresistible (especially with the mid-term elections in the United States just a few months off). But there is also a good chance that there will be progress in the multilateral talks at Stockholm; if some sting can be taken from the fear of a war in Europe, it should be possible for the strategic negotiations to be flexible.

Finding a realistic framework for a year's work at Geneva will nevertheless be difficult. The Soviet Union and the United States, in the weeks before the summit, were talking about schemes for reducing strategic missile forces by about a half, but in terms that make each side's proposals unacceptable to the other. The Soviet plan has the added complication of counting as strategic missiles the French and British nuclear forces, over which the United States has no control. Attempting from Geneva to legislate for these missiles would be fruitless, at least in the immediate future. So the best strategy for the negotiators will be to adopt the goal of a 50 per cent reduction as one for 1987, and, during 1986, to aim at the more modest but attainable goal of putting the existing unratified agreement on strategic arms, SALT II, on a permanent basis that the US Congress will accept. Especially if there is some movement at the Stockholm negotiations, it should also be possible to deal with intermediate missiles by freezing them (with the proviso that the planned US deployment in Western Europe should be completed).

But why such a modest goal? Chiefly because a more ambitious agreement could not possibly be completed in the time available, but also because an agreement to make SALT II stick (it lapses in 1991) would force on the two governments concerned experience of the practical operation of arms control agreements which they at present lack. The idea that there should be regular meetings to discuss what missiles are where, and formal attempts to construct inventories of hostile forces, is unfamiliar. Operating such a system at strategic levels much like the present is a necessary preliminary. Aiming for a comprehensive test-ban now is similarly unrealistic. That is a task for 1988.

The complication of SDI should not be the obstacle public rhetoric suggests. The US administration insists that SDI is a means of making the world safe from the threat of nuclear weapons, the Soviet Union that it is a trick to make the United States alone immune from nuclear retaliation and therefore a provocation. If SDI is technically feasible, both views are tenable. But nobody can tell whether SDI is a realistic development programme or simply a technological mirage. The only reasonable technical prospect is that SDI will spawn, as a minor by-product, a satellite-borne warning system based on infrared detectors. So the Soviet interest would be met by an arrangement that the general character of SDI experiments beyond the atmosphere should be disclosed by a formal agreement to ban the development and testing of anti-satellite weapons and by the clarification and reaffirmation of the Anti-Ballistic Missile treaty. So much could be accomplished by the autumn of this year. □

French research strategy

CNRS embraces plans for research networks

THE French multidisciplinary research council, the Centre National de la Recherche Scientifique (CNRS), is learning to live on a static budget after a rapid period of expansion, its director-general Pierre Papon implied this week. This year's budget of FF8,000 million (roughly £800 million), although up by 8 per cent, will not be enough to cover the rising costs of keeping productively at work the 10,000 of France's most able researchers (10 per cent of whom are foreign) and the 15,000 technicians, engineers and administrators on CNRS's books. But CNRS does not intend to stagnate. Papon says that it will continue to move into new fields of research by adopting new methods of management and organization.

Papon's intentions make an interesting contrast with those of research organizations elsewhere, in Britain for example, similarly constrained by static budgets. Science minister Hubert Curien has effectively ruled out increased concentration of research resources on centres of excellence, saying that would be an unforgivable waste of research talent. This position is supported by Professor Frederick Joliot, science adviser to the prime minister, who argues that the US model for research support entails "excessive" competition, and that the French alternative, based on protected careers in research, allows greater freedom for the pursuit of long-term and speculative projects.

The key word in French science planning now is neither "selectivity" nor "concentration" but "networks", a term once fashionable at the European Commission in Brussels, but introduced in the context of CNRS by M. Curien just before Christmas. The intention is to link existing laboratories and their staffs into new projects, each lasting perhaps only a few years, to share staff and equipment without a concomitant increase of costs.

So Papon and Curien intend that researchers should be more willing to change disciplines and that laboratories should be more open to others, both in France and elsewhere in Europe. Papon, for example, wants to create networks for the study and development of new materials, an approach that contrasts with the decision last year to set up a special laboratory devoted to one of the most fashionable of these, gallium arsenide.

Since 1981, CNRS has been creating new laboratories, usually linked with universities, at the rate of 20 a year, but last year there were only 10 and, during 1986, there will only "a few" additions to the present roster of 750 CNRS laboratories.

Even so, expansion will not be entirely halted. In what may turn out to be a last gesture towards expansionism, CNRS will build a dozen laboratories in the general field of information technology by the end of the decade at Marne la Vallée, a new town still under construction.

Growth will also be sustained by cooperation with regional governments, which now spend up to 3 per cent of their total budgets on research, and with industry. Thus, 1986 will see the beginning of a new metallurgy laboratory in cooperation with the company St Gobain and an inorganic polymer laboratory founded in conjunction with Rhone-Poulenc. But Papon insists that the work at the new laboratories will be basic and strategic — St Gobain, for example, is interested in metallic glasses — and that CNRS will not slip into the role of an applied research agency.

CNRS also intends in 1986 to create several "national laboratories" at which equipment and technicians will be provided only on condition that they are made available for use by other groups for a proportion of the time. An experiment on these lines in engineering science during 1985 is seen as a success and other disciplines are likely to follow in 1986.

During the year ahead, CNRS will have to find FF50 million for projects under the

banner of Eureka, the European collaboration in high technology, out of the 8 per cent budget increase of FF700 million, but even so it will be able to increase spending on medium-scale equipment by 24 per cent, on minicomputers by a third and on large-scale equipment by 13 per cent. Spending on basic research materials will increase by 10 per cent in the life sciences, by 12 per cent in engineering and by 7 per cent overall.

CNRS intends to deal generously with some of its special programmes. Thus the budget of an interdisciplinary programme of materials research begun four years ago will increase by 58 per cent (to FF9.5 million), while there is to be a grand new programme (with a budget of FF11 million in the first year) to bring organic chemists and molecular biologists together.

CNRS will expand by a quarter its department concerned with the application of research, funds allocated to international cooperation are to increase by 13 per cent (international subscriptions are not included) and, in the scientific programme, there is to be more emphasis on the Earth sciences and oceanography (probably at the expense of astronomy); there is, for example, to be a new oceanography laboratory at Toulouse using remote-sensing data, particularly those from the SPOT series of satellites, not to mention a new programme of "cheap" communications research involving neurobiology, cognitive science, linguistics and artificial intelligence. Stasis does not mean to CNRS what it means to other people.

Robert Walgate

Nuclear winter

Soviet researcher's part hyped

VLADIMIR Valentinovich Aleksandrov, the Soviet computer programmer who disappeared in Spain last March, was an "irreplaceable" Soviet scientist whose disappearance put an immediate end to East-West cooperation in the "mathematical modelling of global biosphere processes", in other words, research into nuclear winter, according to the Soviet weekly, *Literaturnaya Gazeta*. After a belated announcement by the Academy of Sciences (his employers) in December, which said little more than that he had disappeared, and a subsequent polemic in *Izvestiya* accusing the Central Intelligence Agency of having kidnapped him, *Literaturnaya Gazeta* devoted a full page to Aleksandrov and "nuclear winter".

By this account, Aleksandrov "is one of the scientists who have made the weightiest contribution to this field of research, thanks to which the consequences of a nuclear war have been placed before the entire world in an entirely new light, even more terrible than before". (He was in fact never more than a junior member of

the Soviet teams to ENUWAR symposia discussing nuclear winter.) In the light of his lack of a doctoral degree, some Westerners have suggested that his place in the delegations may have been dictated by political rather than scientific motives; it is now claimed that he was too busy campaigning about nuclear winter, and that he was due to defend his doctoral thesis on 20 April.

According to his superior at the Computer Centre of the Academy of Sciences, Nikita Moiseev, despite his "comparative youth" (47), his election as a Corresponding Member of the Soviet Academy of Sciences was a foregone conclusion, although he has not been elected *in absentia*.

In one respect the paper is incorrect — international cooperation in "nuclear winter" studies was not "instantly broken" by Aleksandrov's disappearance. An ENUWAR symposium took place in June at the University of Essex, with a Soviet delegation clearly embarrassed by Aleksandrov's absence, to which it referred as a "tragedy".

Vera Rich

German research agency

Ethologist takes the helm

Hamburg

PROFESSOR Hubert Markl, the distinguished ethologist from the University of Konstanz, is at 47 the youngest-ever president of the Deutsche Forschungsgemeinschaft (DFG), West Germany's principal agency for the support of basic research with federal government funds. Markl succeeded Professor Eugen Seibold, the geophysicist, at the beginning of the year. He has been elected as president for a period of three years in the first instance, having served as Seibold's vice-president.

Markl says that the appointment of one as young as he is "an experiment". On the telephone last week, he said that he regretted having to give up his university tasks and his scientific work for science administration, but that he was eager to fall in with the wishes of the members of DFG that he should be their president.

In his own way, Markl is somehow the representative of a new generation of scientists, and is endowed with remarkable optimism, dynamism and farsightedness. He says his most important tasks are "the justification of science to the public" and "the defence of the autonomy of science from which scientists cannot retreat".

Markl has strong views about the development of the university system in West Germany. On the contentious issue of the role of elite universities, Markl says that "one cannot simply copy systems that work well in other countries".

He sees no place in Germany for private elite universities, but instead looks for more competition between universities for good students. This, he believes, could lead to a system in which universities are distinguished by the fields in which they

excel, with one being "top in agriculture, for example, another standing out for its chemistry". In this way, Markl says, "our scientific productivity will be used more effectively".

With the reduction of demand for student places now in prospect on demographic grounds, Markl believes that changes of the kind that he would welcome will come about automatically, especially as universities have more resources to devote to graduate education. He acknowledges that the age structure of university staffs in Germany is at present unsuitable, but hopes that DFG, with its annual budget of DM1,000 million, will be able to provide some relief. He is not especially alarmed at the risk that younger scientists will be sucked away to posts abroad, believes scientific exchange to be an important element of personal development, but says he would act to counter any threat of substantial loss of able people.

Markl has strong views on the social responsibility of science, which he says belongs not to "science as such, but to the individual scientist. Moral precepts never apply to abstract systems, but to the particular people who work for them."

He knows what he is talking about from his eventful life so far. Markl completed his doctorate (in Martin Lindauer's group at Munich) at the early age (by West German standards) of 24, and afterwards spent a year in the United States (at Harvard and the Rockefeller Universities). At the age of 29, then at Frankfurt, he completed his *Habilitation* (the formal qualification to hold a university professorship) as a zoologist, moving in turn to Darmstadt and Konstanz (in 1974).

Markl is likely to enjoy his spell as president of DFG, although his habit of saying what he thinks directly, even bluntly, could cause trouble from time to time. Markl recognizes the danger, saying that the years ahead will give him "a practical training in the applied ethology of science politics" and that "if I choose to walk through fire, I cannot complain of the heat".

Jürgen Neffe

Antarctic treaty

Argentina expands research effort

ARGENTINA is to support its territorial claims in Antarctica by means of a major programme of research and geological exploration. According to the journal *Noticias Argentinas* last month, Argentina may also agree to an international system for the control of Antarctic mineral resources, but this, the paper said, would not prejudice Argentina's "reservations" about its sovereignty rights and maritime jurisdiction.

Argentina is alone among the seven countries that had laid territorial claims to the Antarctic continent before the 1959 Antarctic Treaty in still maintaining the apparatus of statehood there (including passport control). (It was partly the overlapping territorial claims of Britain, Chile and Argentina that led to the formulation of the treaty by which all claims to sovereignty were "frozen" until 1991.) Since the signing of the treaty, mineral wealth in the sector claimed by Argentina has been found to include chromium, nickel, cobalt, copper, gold, silver and iron.

Argentina maintains 16 bases on the Antarctic continent, including six permanent establishments, but only one of these has been engaged on scientific work. One result is that Argentina has carried out less scientific research than the other signatories of the treaty.

During 1985, Argentina sent 33 scientists, mainly geologists, to the Antarctic. In 1986, this number will be increased to 107 "scientists and technicians"; the possi-

bility of opening a further base with a major airstrip is under serious consideration. The new emphasis on science is being presented as a "demilitarization" of Argentine Antarctic policy, although it is officially admitted that this will not mean a reduction in the logistic support provided by the Argentine armed forces. The (military) Argentine involvement in the Antarctic is said to cost some US \$3 million a year. Partial "demilitarization" could make the effort partly self-financing, if Argentina sell equipment and services for Antarctic research and rents out its naval icebreakers for transporting scientific expeditions.

Official Argentine sources explain the need for the new policy by claims that New Zealand's "anti-nuclear" policy could lead to that country being used as a starting point for expeditions sponsored by the socialist bloc. There have also been unofficial indications that Argentina (and also its main Antarctic "rival" Chile) has been perturbed by the new "third world" interest in the Antarctic, demonstrated by the major expeditions by India and China which, it is considered, are operating far outside their own bailiwick.

The Indian position on this issue is, however, explicit; Antarctica is the southern boundary of the Indian Ocean, and may have an important influence on the monsoon and on the climate of India in general. China and Brazil, the two other recent recruits to the Antarctic Treaty, have less specific interests.

Vera Rich

Seibold in Europe

Hamburg

PROFESSOR Eugen Seibold, who retired as president of DFG at the end of 1985, after serving two successive three-year terms, has been for many years one of the chief advocates in West Germany of international collaboration in research. A geologist by training, he was largely responsible for the development of the Geological Institute at the University of Kiel in the 1960s and 1970s.

Since the appointment last year of M. Hubert Curien as a minister in the French government, Seibold has been president of the European Science Foundation. In that role, he says, he will encourage the more rapid completion of the European Geotraverse, the deep seismic sounding of a dog-leg North-South section from Scandinavia to North Africa that is one of the foundation's chief projects.

Jürgen Neffe

Nuclear safety

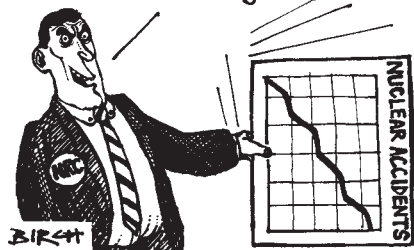
Opposition to relaxing standards

Washington

A ROW is brewing about the official calculations of the amounts of radioactivity likely to be released in nuclear accidents. The Union of Concerned Scientists (UCS) has added its criticisms to the reservations expressed last month by an advisory committee to the Nuclear Regulatory Commission (NRC) about the commission's proposed revision of "source terms" for nuclear power plants. These quantities refer to the probable amount and chemical form of radioactive releases to the environment in the event of a severe accident. UCS says the new source terms "deliberately" omit plausible accident sequences, including some that have actually occurred.

The new source terms are derived in a NRC draft report, NUREG-0956, and are inspired by research in the wake of the accident at Three Mile Island, some of which suggested that releases of radioactivity had been overestimated.

It all depends on how you look at things....



NRC's draft was criticized last month by its own Advisory Committee on Reactor Safeguards, which expressed "reservations" about the document and said the new source terms should "not be given much weight". Containment failure, an important factor in an accident, is treated, in a "rather preliminary fashion".

The NRC draft acknowledges the many uncertainties surrounding its estimates, and admits that the results depend critically on the design of particular plants. But externally initiated events, such as earthquakes, are not considered at all, according to UCS.

Although only five plants are modelled in the draft, NRC says it intends to go ahead and use the new source terms to re-evaluate and revise regulations governing power plant safety, from siting requirements to the design of safety systems. According to UCS, the revisions will in almost every case result in safety requirements being relaxed.

NRC's approach relies heavily on probabilistic risk assessment, which requires analysts to specify all possible routes to failure. Rupture of steam generator tubes, which has led to accidents in the past, is one omission cited by UCS, which has

compiled a lengthy list of similar examples.

The UCS criticism comes at a bad time for NRC, which is investigating several accidents at nuclear plants that, according to Robert Pollard, UCS's nuclear safety expert and a former NRC official, occurred because NRC had not enforced its own requirements. The Toledo Edison Company may be fined \$900,000 because of violations that led to an incident on 9 June last year, at its Davis-Besse plant in Ohio, when both main and auxiliary feed-water were lost. NRC has also upgraded

Neutron research sources

US ambitions for new machine

Washington

COMPETITIVE juices are already rising in US laboratories hoping to capture a new advanced neutron source that they hope will be built during the first half of the next decade. Development funds for the instrument, which is seen as essential if the United States is to catch up with Europe on uses of slow neutrons, will be included in the Department of Energy (DoE)'s research budget for 1987, to be presented by the President to Congress next month.

Several recent studies requested by DoE have emphasized the European lead, which is ascribed largely to the success of the British/French/German collaboration at Institute Laue-Langevin in Grenoble in developing instrumentation for use in cold-neutron experiments. (Little account has yet been taken in the United States of the new British machine, the Neutron Spallation Source, commissioned last year by the UK Science and Engineering Research Council.) There is agreement that the immediate need is for a steady-state source producing up to 10^{16} neutrons $\text{cm}^{-2} \text{s}^{-1}$ (although the project would just get by with half-flux). Design concepts described at a recent workshop in Gaithersburg, Maryland, imply a facility costing about \$300 million with running costs of about \$25 million a year, although part of this could be saved by closing down one of the older sources when the new machine came on-line, according to Dr Donald Stephens, associate director of basic energy sciences at DoE.

Researchers are expecting about \$6 million of development funds for the new machine in 1987. The most popular approach seems to be a 2–300 MW reactor source; Brookhaven National Laboratory is backing a design using small uranium oxide pellets, while Dr Ralph Moon of Oak Ridge National Laboratory wants a reactor with high-density uranium silicide fuel. A design suggested by Argonne

its investigation of an accident in December at the Rancho Seco plant in California in which 450 gallons of contaminated water were spilled in an auxiliary building, leading to a small release of radioactive steam.

NRC says the UCS critique will be considered as a "comment" in its revision of NUREG-0956: a definitive report will be issued later this year. But Pollard doubts whether NRC will take notice "until we get them in court". How soon that will be depends on to what uses NRC puts its new source terms; but if the two do meet in court it will not be for the first time. Last time, UCS challenged NRC's right to restrict public comment on NRC rule-making. It won.

Tim Beardsley

National Laboratory, and disparaged as too preliminary by Moon, has rotating overlapping rings of fuel that would generate high flux densities near their point of contact and cool down during the rest of a revolution.

The chief technical problem foreseen for heavy water-cooled reactors operating at high power densities is in the formation of insulating aluminium oxide on fuel supports. Dr Robert Burke, of DoE's Hanford Engineering Laboratory in Washington state, favours instead a spallation design using a proton accelerator operating at a steady current of 100 mA. Burke says this approach would be more flexible than a reactor (because the protons could be used for other purposes) and that the less onerous safety requirements of an accelerator design mean that siting would be less critical. (He did point out that low electricity costs in the north-west Pacific would make this area attractive.)

The Hanford proposal would cost \$300–450 million, according to Burke (although Moon says that \$800–1,100 million would be nearer the mark), and he claims that the 100 mA proton-beam current required is feasible without extensive research. As a bonus, the Hanford design would allow the source to be upgraded easily should this be thought desirable.

For all the sparring over the design approaches, persuading Congress to build, or even to study, a new facility is not likely to be easy. But DoE's Energy Research Advisory Board has pointed out that existing major US neutron research facilities will be 30 years old by the time a replacement can be built, and the principal use of the new machine — neutron scattering — is the sort of research that can often bring immediate and tangible benefits. That might give it a better chance on Capitol Hill than the much more expensive facilities required for, say, high-energy physics.

Tim Beardsley

US/Canada acid rain

Envoys' compromise in dispute

Washington

A JOINT report on acid rain by US and Canadian special envoys met mixed reactions when it was delivered to the two governments last week. In Canada, there was a cautious welcome for the recognition that acid rain is a serious problem and that scientific knowledge is adequate to assess the effects of control strategies, but disappointment that no specific programme to reduce acid emissions was recommended. In the United States, President Reagan called the report a "step forward" and undertook to study it "carefully", but he avoided endorsing the principal recommendation that the US government should support half the cost of a \$5,000 million programme to demonstrate new coal-burning technologies that minimize acidic emissions.

A Canadian government spokesman said that the report had "brought the acid rain question into the political arena, where it will be solved". Canada, whose tourism and fishing industries stand to lose from damage to lakes and forests, has long held that the scientific evidence justifies measures to reduce emissions from the tall stacks which are the principal sources of long-range pollution. Provincial governments in eastern Canada last year agreed with the national government that they will be responsible for reducing emissions of sulphur dioxide by 50 per cent by 1994. Automobile emissions of nitrogen oxides have also recently been brought into line with the more stringent US standard of 1 gram per mile. The US government, in contrast, has so far refused to acknowledge a link between acid emissions and environmental damage.

Richard Ayers, a lawyer for the Natural Resources Defense Council, says the report is a political compromise which nevertheless allows the utilities to delay action for another five years — well worth \$2,500 million. But a more pessimistic view was offered by Adele Hurley of the Canadian Coalition on Acid Rain, who does "not expect to see the money".

The two special envoys, Drew Lewis (United States) and William Davis (Canada) were appointed last March to try to avoid a political impasse over the issue. There is no question that far more acid emissions find their way from the United States to Canada than in the opposite direction. The envoys say in their report that "it is very clear there is a solid link" between acid precipitation and emissions of acidic gases. Although clean-air legislation enacted in the United States in the early 1970s has improved ambient concentrations, long-distance transport (a result of tall smokestacks) has not improved.

The most widely used method of reduc-

ing emissions from existing plants, the scrubbing of flue gases, is expensive and produces enough waste to be a problem in its own right. New technologies that burn coal with increased efficiency and remove sulphur during combustion rather than afterwards, such as combined cycle gasification and fluidization bed combustion, offer a cheaper way to remove pollutants, but utilities say that they cannot be expected (because their tariffs are regulated) to risk huge sums on unproven technologies.

Meanwhile, the electricity utilities welcome the contention of the Lewis/Davis report that present scientific evidence is insufficient to justify "punitive control". Industry spokesmen point out that utilities have already voluntarily spent \$500 million on research into clean coal technology over the past 5 years, to be followed by another \$500 million over the next 5 years.

The utilities' support for the envoys' recommendations is not altogether surprising because, if adopted, the recommendations would ensure that the US government pays half the cost of technology demonstration projects. The US Congress at the end of last year put \$400 million into clean coal technology, but the industry has undertaken to spend up to \$2,100 million if it gets a promise of support.

Some who have studied acid rain are less happy about the Lewis/Davis assessment that technology demonstrations are what is called for. David Schindler, who

chaired a study on acid rain for the National Academy of Sciences in 1981, said he was "disappointed" that immediate control measures were not recommended and that the report concealed its true purpose of delaying action. Schindler said the evidence implicating sulphur dioxide in the acidification of lakes was "clearly strong enough" to support immediate action and stronger than the evidence on which the use of phosphorus in detergents was banned in the 1970s because of its effects on the Great Lakes.

Jack Calvert, of the National Center for Atmospheric Research, who chaired another National Academy study, said he had hoped the Lewis/Davis report would recommend a reduction in emissions from some of the major polluters in Pennsylvania, West Virginia or Ohio, so that actual effects could be observed and an informed decision taken on whether further controls are likely to be effective. Calvert applauded the report's support for greater efforts to monitor acid deposition, especially dry deposition, which is still not monitored systematically.

Whether the special envoys' work will lead to an easing of tensions between the United States and Canada over acid rain will now depend on the US administration's response to their recommendations, which include the setting up of a joint committee with a US cabinet-level representative.

The issue will be a "priority topic" at a meeting to be held in Washington in March between President Reagan and Canadian Prime Minister Brian Mulroney.

Tim Beardsley

Fusion research

US, Europe, Japan collaborate

IN an agreement signed yesterday (15 January), the United States, Europe and Japan, traditionally rivals in the search for economic energy from nuclear fusion, formalized a programme of enhanced cooperation in this direction. The agreement, signed at the Max-Planck-Institut für Plasmaphysik at Garching, Munich, commits scientists from the three programmes to hold regular trilateral workshops, to exchange basic data and experimental plans, and also personnel and equipment.

All three programmes are tokamak machines in which ionized hydrogen trapped in a toroidal container by magnetic fields may be heated to temperatures and densities sufficient for fusion reactions to occur. Basic heating is provided by passing a current through the torus, treating the plasma as a secondary winding in a transformer. But supplementary heating is required and, at all three tokamaks, is provided by a combination of high intensity radio waves generated from antennae in the torus walls and the injection of neutral

atoms at high energy.

Present interest at all three tokamaks is in the development of the supplementary heating systems; the problems of magnetic confinement and heating seem to have been solved. The heating system for Japan's JT-60 tokamak is expected to be working by next December, while at the Joint European Torus (JET) at Culham in Britain, the first stage of radio heating has been tested and neutral beam injection is about to be installed. Princeton's Tokamak Fusion Test Reactor (TFTR) is to embark this week on a lengthy experimental run.

According to Donald Grove of TFTR, the new agreement should make the three programmes more productive. "From now until the international fusion conference in Kyoto next November, the three tokamaks are on essentially the same footing", he says. It is hoped TFTR will achieve break-even in energy production, by means of deuterium fusion, by 1987.

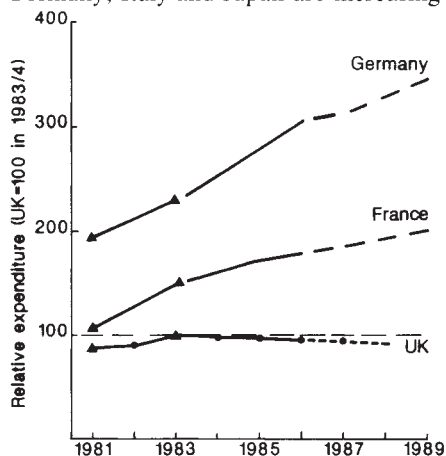
Philip Campbell

British science

Researchers find a voice

BRITISH scientists, traditionally quiet and uncomplaining, have expressed their anger, in the form of Save British Science (SBS), an "appeal to public and Parliament to call on the Government to reverse the decline in support for scientific research", which was launched in London this week. Four Members of Parliament (MPs) are proposing a debate in the House of Commons, and will ask the government to spend £100 million this year on basic science research — enough to keep Britain level with other European countries.

Professor Denis Noble, one of the scientists launching the appeal, said this week that "a few years ago it would have been unthinkable that someone in my position would speak out publicly" about the state of science spending in Britain. The decline in real terms (see graph) at a time when France, West Germany, Italy and Japan are increasing



Government funding of basic research in the United Kingdom, France and West Germany. Triangles, data from the *Annual Review of Government Funded R & D 1984 and 1985*. Dotted lines, policies based on recent ministerial reports.

research spending, means that "selectivity is becoming a polite word for retreat, which is now in danger of becoming a debacle", according to Dr J. Mulvey, secretary of SBS. Noble described examples of scientists leaving to work abroad, obsolete equipment and occasions where scientists who had written successful grant proposals did not receive enough money for their own salaries, but only enough for the project itself.

SBS has collected more than 1,500 contributions since it was formed last November, and claims that grassroots support is "snowballing". But, apart from an advertisement in *The Times* about the predicament of British science, no coherent policy has yet emerged. Others have already pointed out the dangers of failing adequately to support basic science research to the government, with little success. The group must turn into efficient lobbyists for parliamentary support, and, as suggested by the MPs present at its

launch, be both apolitical and efficient in communicating with the Government. SBS should have a formal membership so that it can genuinely reflect the wishes of its supporters and become an effective force in popularizing the need for research. It wants to obtain more support for research from industry, and thinks that government-induced collaborations between industry and academics are unlikely to produce more money for basic research

Refusnik scientists

Dissenting dissidents agree

LEARNED societies campaigning for oppressed scientists in the Soviet Union should avoid mentioning "human rights" in letters to the Soviet scientific establishment, Dr Alexander Voronel of Tel Aviv University told a seminar in Oxford this week. The seminar, for representatives of national academies of science and learned societies, had met to discuss what such bodies can do to help Jewish *refusnik* scientists. Although there were some vehement objections to Dr Voronel's view, notably from a delegate from the United States who described himself as an "academic lawyer", Dr Voronel's record as founder and first convener of the original Moscow Sunday seminar for *refusniks* gave his views considerable weight.

Dr Voronel's contention is that any mention of "human rights" is an automatic irritant to Soviet officialdom, which counters by alleging that the West is trying to interfere in internal Soviet policy. One should rather campaign, he said, for Soviet scientists to be allowed to exercise their professional rights of appropriate employment, unrestricted professional correspondence, participation in scientific gatherings both at home and abroad, and — the particular issues in the case of *refusniks* — the right to emigrate. This "right", incidentally, is now of paramount importance to the *refusniks*; a message to the seminar from three leading *refusnik* scientists, Victor Brailovskii, Mark Freidlin and Aleksandr Ioffe, urged all well-wishers "in view of the extreme urgency of our problem, [to] put aside all other aspects of our being here, and the difficulties encountered in our daily lives" and campaign only on the emigration issue.

Mr Fabian Kolker, the US philanthropist, who two years ago founded a scholarship for ex-*refusniks* at Tel Aviv University, summed up the views of both Voronel and the *refusniks'* letter. "Our goal should be success in unlocking the gates of the Soviet Union", he said, "and we must therefore use language that the Soviets will accept."

whereas a tax-credit system for industries that fund this type of research is more likely to achieve the aim.

In Parliament, SBS could investigate the setting up of a body like the US Office of Technical Assessment, to advise MPs on scientific issue and to contribute to informed debate and wider general acceptance of the importance of science. A policy must also be formulated on whether or not Britain should have a representative at cabinet level specifically for science, or how seriously to take the Prime Minister's contention that she represents the interests of scientists.

Maxine Clarke

How to put more leverage behind such "language" was the main point of discussion. The chairman, Dr Inga Fischer-Hjalmars of the University of Stockholm, presented an 11-point programme of suggested action, particularly slanted towards making intervention by Western colleagues a "professional" and not merely a humanitarian matter. Thus *refusniks* can be elected to the appropriate learned societies, or asked to join editorial boards of scientific journals, giving such societies or boards a valid interest in the welfare of and continuous contacts with their fellow members. (A suggestion from the floor extended this to asking *refusniks* to review books, which would have the additional bonus of supplying them with works in their field.)

If learned societies are to help, however, they must concern themselves, the consensus of the seminar agreed, not just with *refusniks* but with the whole range of breaches of human rights under all forms of repressive regime. This has always been a divisive issue, with Israeli scientists maintaining that one must not confuse *refusniks* and dissidents, while their foreign colleagues pointed out that it was impossible to ask a fellow faculty member to sign a petition on behalf of Brailovskii or Shcharanskii, and they refuse to sign the petition he was organizing on behalf of oppressed scientists in Turkey or South Africa. This time, however, there was little contention. It is now accepted, even by the Israelis, that learned societies and academics have human rights committees that campaign across the political board for scientists victimized for their convictions. Nevertheless, as Dr Kristoffer Gjotterud of Oslo pointed out, the distinction is a real one, and in campaigning for Soviet scientists in trouble, he stressed that one should never confuse the dissidents who wish to change what they see as inherent injustices in the Soviet system and the Jewish *refusniks* who are penalized for wishing to leave the Soviet Union altogether.

Vera Rich

Viral cancers

AIDS virus at the centre

Los Trois Ilets, Martinique

"RESEARCH in viral-associated cancers has come of age", said Peter Fischinger, deputy director of the National Cancer Institute (NCI) at a three-day conference here co-sponsored by the National Institutes of Health (NIH) and the Association pour la Recherche sur le Cancer (ARC).

Initial enthusiasm for the possible role of viruses in cancer accompanied the discovery in the 1960s that Epstein-Barr virus is associated with Burkitt's lymphoma. But that enthusiasm abated when little evidence was found to implicate other viruses with human cancers.

Today, the enthusiasm has returned. Nasopharyngeal cancer, a much more significant worldwide health problem than Burkitt's lymphoma, has also been linked to Epstein-Barr virus. Hepatitis-B virus DNA has been found in hepatocellular carcinomas. Different variations of human papilloma virus have been identified in 80 per cent of human papillomas, and, according to Harald zur Hausen of the Deutsches Krebsforschungszentrum at Heidelberg, it is likely that some form of human papilloma virus will be found in all forms of human cervical cancer.

Probably the most significant reinforcement of interest in cancers associated with virus comes from the discovery that a

human retrovirus, HTLV-I, causes adult T-cell leukaemia, and that another human retrovirus, HTLV-III/LAV, causes acquired immune deficiency syndrome (AIDS).

In 1985, NCI spent \$135 million on viral cancer research, approximately 12 per cent of its budget. Whether that level of support will continue is unknown. Both NCI director Vincent DeVita and NIH director James Wyngaarden were forced to bow out of this conference at the last minute to devise a new budget for NIH under the guidelines set forth in the recently passed Gramm-Rudman legislation (see *Nature*, 318, 495; 1985).

Although HTLV-III/LAV has not been directly associated with cancer, the explanation may be that patients are killed by opportunistic infections seen in AIDS before cancers can develop. According to Max Essex of the Harvard School of Public Health, many more cats die from opportunistic infection associated with feline leukaemia virus than die from leukaemia.

A possible human parallel is seen in Japan; Isao Miyoshi of the Kochi Medical School reports seeing infections similar to those seen in AIDS patients in individuals infected with HTLV-I. There may be more of these patients, but so far no one has thought to look for them. Ironically, it may yet turn out that HTLV-III/LAV will cause cancer if AIDS patients can be kept alive long enough to exhibit it.

The explosion in AIDS research has produced several interesting and potentially important discoveries. Evidence for heterosexual transmission of HTLV-III/LAV continues to grow. In 1979, 12 per cent of AIDS patients in Haiti were female; now, the number has grown to 30 per cent, and is expected to grow to 40 per cent in the next two years.

Conflict continues about what constitutes an effective therapy for AIDS. Research at the Pasteur Institute in Paris has focused on HPA-23, a compound that can block HTLV-III/LAV production *in vitro* and apparently *in vivo*. Sam Brower of NCI believes that a therapy that blocks viral replication is insufficient, and that the need is for an agent that will restore the helper/inducer T cells destroyed by HTLV-III/LAV infection. One such compound appears to be 3' azido-3'-deoxythymidine (AZT), now in clinical trials at NIH. A major advantage of AZT is it can be taken orally, an important consideration as it will have to be taken for life.

There are also indications that a new generation of assays for HTLV-III/LAV is not too far off. An important first step achieved in a preliminary form by Biotech Research Laboratories in Rockville,

Maryland, is the purification of viral coat proteins, which may also play an important role in vaccine development.

There was some tension at the meeting between NCI and Pasteur Institute researchers over the recent lawsuit by the French over the granting of a patent for an HTLV-III/LAV antibody test (see *Nature* 318, 595; 1985). Both Luc Montagnier, the Pasteur researcher who first reported the association of a retrovirus with AIDS, and Robert Gallo, the NCI researcher whose work led to the US patent, attended the meeting. Nomenclature became somewhat confusing as the virus was alternately referred to as "HTLV-III/LAV", "LAV/HTLV-III", "LAV" or "HTLV-III" depending on the nationality and affiliation of the speaker. Future collaborations are not likely to suffer because of the current contretemps.

But Gallo complains of one unpleasant side effect of the lawsuit; he has been ordered to make photocopies of all notes and data relating to his work by US government lawyers acting in the lawsuit; this, Gallo maintains, will set back his research schedule.

ARC, the French co-sponsors of the meeting, have taken an active interest in viral associated cancers, especially in their relevance to health issues in underdeveloped countries. They also pride themselves on bringing together researchers around the world to tackle problems of mutual interest. In March 1984, a conference in Marseilles on cancers of the Mediterranean included Arab and Israeli researchers.

Says Gerard Milhaud, a member of ARC's scientific advisory board, "We got the Israelis and the Arabs together for a meeting. Bringing Montagnier and Gallo together was easy." **Joseph Palca**

Infection mechanism?

NCI's Robert Gallo now believes that transmission of HTLV-III/LAV infection is through cells, not virus particles. His scenario for spread of infection suggests that when a donor cell is recognized by a recipient cell, there is a mixed lymphocyte reaction. This in turn releases lymphokines which activate the initiating sequences of the viral genes, including the transactivating transcription gene (*tat*), which in turn causes viral expression in the newly infected host. This mechanism would explain why HTLV-III/LAV is not highly contagious, requiring some transmission of body fluid containing lymphocytes.

Gallo also reported five new cases of HTLV-II infection, and suggested there is some evidence that this virus is also spreading.

Although evidence for brain infection by HTLV-III/LAV is now undisputed, the exact target for infection is still in doubt. Luc Montagnier of the Pasteur Institute believes glial cells are infected directly, and has *in vitro* evidence to support this. Gallo, however, believes that brain involvement comes from monocyte macrophages migrating to the brain, possibly turning into microglial cells.

Joseph Palca

Polish astronomers

THREE Polish astronomers from the Copernicus Astronomical Centre in Torun are expected to go on trial this week for possession of an illegal radio transmitter. Drs Jan Hanasz, Zygmunt Turlo and Leszek Zaleski, all members of the Polish Astronomical Society, were arrested on 26 September, together with Mr Piotr Lukaszewski, a technician from a local factory, charged with an offence against public order and possessing the transmitter.

They had used it to put out a television broadcast calling for a boycott of the forthcoming general elections, which they categorized as "undemocratic".

The astronomers last weekend called on Western colleagues to take note of their case, urging that, under prevailing conditions in Poland, the normal methods of making one's viewpoint felt were not available.

Vera Rich

British veterinarians

Pressures for collaboration

THE six British universities providing courses in veterinary medicine are in the throes of deciding how to meet the demand that they should teach fewer students and, at the same time, cost the central budget less. At a meeting on Wednesday this week, heads of the six schools and officials of the Royal College of Veterinary Surgeons will have tried to put a plan together. The University Grants Committee (UGC) is looking for a firm proposal by 28 January, and will otherwise draw up a scheme of its own.

The six British veterinary schools (at Bristol, Cambridge, Edinburgh, Glasgow, Liverpool and London) are all having to adjust to smaller student numbers in the wake of the third quinquennial report on veterinary manpower, commissioned by the Ministry of Agriculture, Fisheries and Food, which recommended a reduction of 10 per cent in the numbers of students recruited into veterinary medicine, from a total of 335 a year to just 300. These reductions will take effect from the beginning of the next academic year.

Pressure from UGC arises from the high cost of teaching veterinarians in Britain. Courses last five years, following the pattern of medical education, with three "undergraduate" years of biological science and two years of clinical study, including a compulsory 26 weeks of on-the-job training. Professor E.J.L. Soulsby (Cambridge), last year's president of the Royal College and thus the initiator of the present reform, points out that the teaching of veterinarians and physicians differs in that part of the cost of medical teaching is provided indirectly by the public health services, which keep the teaching hospitals full of patients and provide for their care. Many veterinary schools have to rely on private practitioners to supply sick animals, which must then be treated and cared for out of university funds.

Several alternative proposals for reorganization are being considered, among them a plan to allow the six schools to continue in their slimmed-down form for the first three years of the standard course, but to concentrate resources required for the succeeding years at different schools, with the understanding that students would move from one place to another.

Professor I.A. Silver (Bristol) says that if the six schools have not agreed on a plan this week, it will fall to the college (of which he is president this year) to put forward a scheme acceptable to UGC. He is heartened that UGC says it will pay close attention to the quality and the amount of research in deciding which schools should be favoured.

Not all veterinarians are reconciled to

the conclusion that student numbers should fall by 10 per cent. Professor A. Iggo (Edinburgh) points to the potential value of novel biological techniques in the future treatment and management of domestic animals and thinks that the manpower estimates may take too little

DNA fingerprinting

DNA probes control immigration

ONE of the first uses of the technique developed at the University of Leicester by Dr Alec Jeffreys for telling the genetic identity of individuals by the use of gene probes is likely to be in the control of immigration into Britain. But the Foreign Office denied last week reports from Dacca, Bangladesh, attributing to Mr Timothy Eggar, Under-Secretary of State, the announcement of the introduction of an operational scheme by April this year.

The issue is potentially contentious between Britain and Bangladesh, and between the British government and its immigrant population, because of the requirements of the now strict immigration law distinguishing between blood relatives of different affinity. What appears to have happened so far is that the Home Office and the Foreign Office are jointly sponsoring a pilot study at Leicester, but have not yet decided whether to mount an operational scheme for verifying claims of kinship.

Jeffreys says, however, that the technique has already been used successfully in a number of disputed immigration cases, the first of them on the initiative of a solicitor (attorney) acting for the father of an immigrant boy whose kinship had been disputed. The solicitor had been alerted to the usefulness of the techniques by publicity following the publication of a paper last March (Jeffreys, A.L., Wilson, V. & Thein, S.L., *Nature* **314**, 67; 1985; see also *News and Views* **318**, 507; 1985). The first analysis proved the claim of kinship, but the Home Office withdrew from the case so that there is as yet no legal precedent in Britain for the use of evidence of this kind.

Jeffreys says that the use of DNA fingerprinting in immigration cases is likely to be more complicated than in common paternity cases because disputed kinships may centre on suspicions that an adult is seeking to pass off a nephew as a son. Because of the high frequency of common genes among siblings, it is therefore necessary to test for a greater number of restriction polymorphisms. Even so, he estimates that it should be possible in routine cases to attain a precision of more than 99 per cent in cases of disputed

account of research needs. He reckons that a third of the income of his own department may arise from research grants and contracts.

Most British schools appear, however, to accept that the overseas demand for veterinarians has shrunk substantially, even though there is still a buoyant demand from overseas schools from students with a first degree seeking graduate training and experience. □

kinship.

The Home Office emphasizes that even if the technique is introduced on a regular basis in considering applications for immigration to Britain, this will be offered only on a voluntary basis. (Nothing is said of the fate of those who decline a blood sample.)

Whatever happens to the immigration project, the forensic use of the technique is likely to be widespread. The commercialization of Jeffrey's version of DNA fingerprinting is being exploited by the Lister Institute. □

Romanian mathematics

Absent guest at Princeton

DR Radu Rosu, a Romanian mathematician invited to the Princeton Institute of Advanced Studies, last week began a 12-day fast to mark the beginning of the second term in which he has not been able to take up his invitation. Rosu, who was due to begin work at Princeton last September, claims that his application for a passport is blocked by the Communist Party organization of the University of Bucharest.

Dr Rosu, a geometer, began his research career at the Institute of Mathematics of the Romanian Academy of Sciences. When the academy was stripped of its institutes in the mid-1970s, and pure mathematics was virtually eliminated on utilitarian grounds, Rosu became a programmer at the computer centre of the Research Institute for the Energy Network.

In 1979, he transferred to the Institute of Mathematics at the university, where, in 1983, he successfully defended his doctoral thesis. The invitation to Princeton seems to have been inspired by a paper with Soviet mathematician I.M. Gel'fand.

Dr Rosu is not the only Romanian mathematician resisting his government's repressive policy. Last week, Dr Mihai Botez gave an interview on French television drawing attention to (and defying), the ban on unauthorized contacts with foreigners.

Vera Rich

Division of Japanese psychiatry

SIR—As the visiting scientist who drew the attention of *Nature* to the events surrounding the seventh congress of the Japanese Society of Biological Psychiatry in April 1985 (see *Nature* **315**, 361; 1985), I was interested to read the contribution of Dr K Moriyama (*Nature* **318**, 308; 1985) to the debate. As I understood the situation (from psychiatrists attending the meeting in Nagoya), it was indeed threats from the group with which Dr Moriyama is associated that led to the abandonment of the seventh congress, a satellite symposium on schizophrenia to be held in Gifu and then the *ad hoc* meeting organized in a hotel in Nagoya.

Two issues have emerged. Undoubtedly there are aspects of the Japanese mental health care system that are at variance with practice in other advanced nations. The report of the fact-finding mission of the International Commission of Jurists and International Commission of Health Professionals (see *Nature* **316**, 571; 1985) highlights a number of problems. One would expect that these would be the subject of discussion within and outside the psychiatric community in Japan and that someone of Dr Moriyama's declared interests would be working within a professional and political context (for example through a body such as MIND) for legal changes.

The second issue relates to freedom of exchange of scientific information. Dr Moriyama writes of "protest action taken against the seventh congress". Undoubtedly he and his followers have created a situation in which organizers feel obliged to cancel meetings rather than face confrontation. I understood from the psychiatrists to whom I spoke at the aborted meeting in Nagoya that previous "protest action" had ended in physical violence with the possibility of serious injury to meeting participants. A visitor to Japan has difficulty in understanding how Dr Moriyama and his colleagues justify adopting tactics that presumably are illegal, and finds it equally perplexing that the congress organizers did not consider the protection of the law adequate to allow the congress to continue.

Dr Moriyama alleges that the papers to be presented at the seventh congress reveal "evidence of many unethical experiments". I have a copy of the volume of 114 numbered abstracts that were to have been presented. These are mostly in Japanese, although the sections that are not indicate that the topics include many that were discussed at the recent World Congress of Biological Psychiatry in Philadelphia. I invite Dr Moriyama to indicate which abstracts he considers raise

particular ethical issues. I will then arrange for these abstracts to be translated and submitted to a United Kingdom ethical committee (see for example Denham *et al.* 'Work of a district ethical committee', *Br. Med. J.* **ii**, 1042–1045; 1979) or to a comparable American institutional review board. It will then be possible to assess whether there is independent support for the views expressed by Dr Moriyama.

T.J. CROW

*Clinical Research Centre,
Division of Psychiatry,
Watford Road,
Harrow, Middlesex HA1 3UJ, UK*

Animal intelligence

SIR—I should like to add to J.H. Fremlim's letter on animal intelligence (*Nature* **316**, 760; 1985). Human intelligence is not only quantitatively but also qualitatively different from animal intelligence because it recognizes time as a dimension. Man knows that there is a future in addition to the past and the present.

The awareness that a future exists gives man the incentive to plan, and to ensure that the plan succeeds. He therefore prepares weapons for the hunt and containers for the foraging expedition. This culminates in the discovery of agriculture, in the knowledge that seed planted in the spring will bring in a rich harvest in the fall, enabling him to survive the winter in comfort.

ANDREJS BAIDINS

*1104 Windon Drive,
Wilmington, Delaware 19803, USA*

Aleksandrov

SIR—Vera Rich writes (*Nature* **318**, 592; 1985) of "rumours circulating in the West" that Academician Vladimir Aleksandrov, who disappeared in Madrid last April, had been intending to defect, but was prevented by Soviet security men.

We no longer live in a time when every member of a Soviet delegation abroad is shadowed by another. Vladimir Aleksandrov, like many of his compatriots travelling in the West, had countless opportunities to defect. At a conference in Helsinki in 1984 I met him early one morning jogging in some woods — with nobody on his heels. And Western embassies in many capitals would surely be glad to give asylum to a leading Soviet mathematician and computer scientist.

It is also simply speculation to suppose that Soviet official silence on Academician Aleksandrov's disappearance indicates sinister goings-on. It is not Soviet policy to utter publicly on events that are

bad news, unless it is to prevent their recurrence. Of course these rumours may be right, but where is the evidence?

STEWART BRITTEN

*9d Stanhope Road,
London N6 5NE, UK*

Creationism

SIR—In his Commentary on the unreasonable claims of scientific creationists, Michael Cavanaugh (*Nature* **315**, 185; 1985) seems to ignore the boundaries in biology that separate some observations from their causes and which are the basis for appealing to history, whether traditional or conjectured. Scientists have deciphered the genetic code, but chemical genetics has not supplied a theory for its origin. The stability of nuclear DNA has frustrated understanding how and from where organisms might acquire the enormous amount of genetic information needed to leap the persisting gaps in the fossil record — and on so grand a scale to bring whole populations into existence suddenly and worldwide. Cosmologists, too, recognize that the equations of the big bang theory are not capable of describing conditions before the big bang.

To assert in the absence of scientific theory that the origin of the genetic code and acquisition of new genetic information by organisms nevertheless occurred spontaneously is clearly a matter of faith. Evolutionists and creationists are therefore epistemological equals who are free to choose which history provides more meaning for life.

Creationists are indebted to evolutionists for having put the history of *Genesis 1* to the test. Practice of the scientific method has established boundary conditions concerning origins that are evidence for a creator. Along with evolutionists, creationists predict that these boundaries one day will open to rational thought.

D.H. KOOBS

*Box 543, Loma Linda,
California 92354, USA*

Is anyone there?

SIR—There must surely be *something* magical about the "magic frequencies" at which, you report (*Nature* **317**, 275; 1985), yet another project is searching for extraterrestrial signals. But the magic is that anyone should waste their time, and our money, searching the bands where interference and attenuation of any such signals is at a maximum.

Any rational ET communication engineer would transmit at frequencies where these are at a minimum. Is anyone looking in the right place?

M.H. MACDONALD

*8 Wilton Street,
Brighouse, Yorkshire, UK*

Newtonian gravitation corrected

Suspicion that the inverse square law for the gravitational force between two objects may be incorrect at short distances appear to have been confirmed. Better measurements are required.

NEWTON'S inverse square law of gravitation may, after all, be incorrect. That is the strange and somewhat mystifying conclusion of the latest investigation of the value of the universal constant of gravitation (G). For although there have been many recent suggestions, theoretical and otherwise, that the strength of the gravitational coupling between massive objects may be a function of distance at small distances, the surprise is that it has now been established by a re-examination of the results of the Eötvös measurements, widely accepted for more than half a century as a proof that gravitational interactions depend only on the mass of the objects involved.

The new development (*Phys. Rev. Lett.* **56**, 3; 1986) is the work of Ephraim Fischbach of the University of Washington, St Louis, Daniel Sudarsky, Aaron Szafer and Carrick Talmadge of Purdue University and S.H. Aranson of the Brookhaven National Laboratory. The essence of their conclusion is that G , the 'constant' in Newton's inverse square law, is in reality a function of the distance separating interacting masses, and is less by a few parts in a thousand at distances of the order of 200 m.

The reasons why G may not be constant have been widely canvassed, and follow naturally from the now common assumption that the four natural forces (gravitation, electromagnetism, the weak and the strong nuclear interaction) are, at some level of abstraction, aspects of one general interaction. Field theorists who have set their sights on the 'grand unification' of the elementary forces have long anticipated this conclusion. In such a scheme, the gravitational interaction between two masses is mostly explained by the exchange of hypothetical field particles ('gravitons'), just as the electromagnetic interaction is accounted for by the exchange of photons between interacting electric charges, but grand unification allows other field particles to chip in, modifying Newton's inverse square law.

The expectations have been nicely put by G.W. Gibbons and B.F. Whiting (*Nature* **291**, 636; 1981) for example, who argued for departures from Newton's inverse square law represented by a G which is a function of the distance r between interacting masses with the form $G = G_0 (1 + \alpha e^{-r/\lambda})$ where G_0 is the value of G at large distances; α and λ are constants, the second of which is deter-

mined by the mass of the extra field particle mediating the gravitational interaction and provides a scale of lengths at which departures from the inverse square law should be apparent.

Geophysical data provide supporting evidence. If G varies with distance, and if the characteristic length scale is of the order of metres, the acceleration of a satellite in orbit about the Earth should differ from that expected from the measured acceleration of a mass near the surface of the Earth and from a direct measurement of the force between two masses separated by a few centimetres, as in the measurements of Cavendish a century ago.

Geophysics should yield information on departures from Newton's law a little more directly, as F.D. Stacey and G.J. Tuck have clearly argued (*Nature* **292**, 230; 1981). They showed that it should be possible to use measurements of gravitational acceleration ('little g ') in boreholes in the Earth's crust, together with measurements of the density of the rock strata they traverse, to estimate the possibility that G varies with distance. Two years ago (S.C. Holding and C.J. Tuck, *Nature* **307**, 714; 1984), results of gravity measurements in two mineshafts at the Mount Isa mine in Australia were held to be consistent with a variable G , although everybody seems to agree that measurements in mineshafts, especially in metalliferous regions, are far from ideal. Stacey seems to have been the first to press for measurements of g as a function of depth in the deep oceans, where inhomogeneity and uncertainty about the density of the surrounding material would be less serious complications.

Fischbach *et al.* take a different tack, going back to the classical measurements in which Eötvös sought to demonstrate that the gravitational attraction between two objects is simply a function of the product of their masses (and the distance between them). Several pairs of different materials were evaluated against each other — water, copper sulphate, platinum and even tallow against copper, for example. Fischbach *et al.* rightly say that the measurements have been taken as a proof that the mass, and not its constitution, is all that matters.

Their own starting point is more complicated. What if departures from newtonian gravitation are brought about by the exchange of massive photons such

as might be associated with the non-conservation of charge and parity (CP -non-conservation, as it was called by Lee and Yang in the 1950s)? Then the extra force should be repulsive, and should be determined not by the masses of the gravitationally interacting objects but by their baryon numbers (or, more accurately, by their hypercharges).

But surely the baryon number of one gram of platinum will be the same as the baryon number of a gram of copper sulphate, and simply determined by the number of nucleons it contains? This is the simple explanation. But allowing both for the contribution of electron masses to the total mass of a material, and for nuclear packing fractions, it turns out that there are more baryons per unit mass in the middle of the periodic table (around Fe) than at either end.

All this is beyond dispute. The surprise is that Fischbach *et al.* have been able to find in the original Eötvös measurements evidence of departures from newtonian behaviour that correlate strikingly with the specific baryon number (the number of baryons per unit mass). Naturally, they go on to plead that the experiments should be repeated, but they also go on to say that their conclusion is sustained by measurements of anomalous properties of K mesons (kaons) at high energy.

What is to be made of the conclusion? The claim in the *New York Times* last week that Fischbach *et al.* have discovered a 'fifth force' is a little strong; if there are measurable departures from newtonian gravitation, those described are much what would be expected. That there is a significant difference between laboratory and geophysical measures of G seems now also well attested. The fact that the scale length calculated for the departure from the inverse square law is of the order of 200 m is consistent with that, but should suggest many novel ways of designing critical tests of the gravitational law now put forward.

That, in the end, is how the matter will have to be settled, for it is unthinkable that such an important issue as the behaviour of newtonian gravitation at short distances could for long be allowed to rest on a series of experiments carried out more than 60 years ago. But Fischbach *et al.* have provided an incentive for the design of better measurements by showing what kind of irregularity it will be sensible to look for.

John Maddox

Space physics

Does Uranus have a magnetic field?

from A.J. Dessler

IN JUST a matter of days, Voyager 2, now nearly 8.4 years into its grand tour of the Solar System, will sweep past Uranus. The prospect has inspired discussion of the strength of the planet's magnetic field, published estimates of which at cloudtop level range from tens of gauss down to zero. This fascinating state of confusion is the result of recent data from Voyager which do not confirm past observations of Uranus from Earth orbit.

In 1982, John Clarke, using the Earth-orbiting International Ultraviolet Explorer (IUE), reported that Uranus is a bright source of ultraviolet light¹. Clarke's interpretation of these observations, which was immediately and widely accepted,

was that the ultraviolet emission from Uranus is radiating from a bright aurora, which would be caused by the bombardment of the upper atmosphere by energetic electrons or ions. It is these and subsequent IUE observations, plus the specific hypothesis of an aurora on Uranus, that led to the estimates of a strong planetary field. To extract power either from the solar wind or from the rotation of the planet so as to drive an aurora that can produce the observed 10^{11} watts of ultraviolet emission, the planetary magnetic field has to create a magnetosphere large enough to engage the power source. Because of our experience with aurora on Earth, Jupiter and Saturn, quantitative

models are at hand. Application of these models to Uranus results in a most probable value of the magnetic field at cloud-top level of 4–13 gauss, with 0.6 gauss as the minimum value².

Until a few months ago, Uranus was, in a sense, an astrophysical object. It can be seen well enough from Earth (and from Earth orbit) to learn a few facts and to develop some nice stories. Because there are few facts, simple ideas tend to fit them and they become a consensus rather quickly. Uncertainty and debate have been rare. But we are now witnessing a transition of Uranus from an astrophysical object to a space-physics object; facts provided by Voyager are about to overwhelm the theories. The impending encounter has made the theoretical community both humble and open-minded.

The problem is that if the IUE-detected ultraviolet light is indeed from an aurora created inside an enormous magnetosphere, then our experience with Earth, Jupiter and Saturn would lead us to expect Uranus to be a significant source of radio noise. Consequently, it was predicted that Voyager's first detection of magnetospheric radio emission would occur early last year and no later than August^{2,3}. The failure of Voyager to pick up any radio signals from Uranus so far suggests that the ultraviolet emission is not generated by an aurora and that a strong (or even a weak) planetary magnetic field may not exist. Alternative explanations for the ultraviolet emission are therefore being put forward.

Shemansky, in the January issue of *Geophysical Research Letters*, and Ip, in next week's *Nature*, agree that the mysterious mechanism that produces bright ultraviolet glow from the sunlit hemispheres of Jupiter and Saturn also operates at Uranus. This glow is not auroral in nature — it occurs at lower and mid-latitudes, stops suddenly in the absence of sunlight and is clearly a separate phenomenon from the standard high-altitude aurora. Shemansky further suggests that Uranus has a negligible planetary magnetic field, while Ip argues that the strength of the magnetic field is immaterial, but it could be quite weak.

The lack of the expected radio emission has inspired yet another argument for a weak magnetic field. Curtis and Mess⁴ note that the blackness of the rings and satellites of Uranus could be explained if there were no magnetosphere to shield them from the solar wind. Particle bombardment by the solar wind would remove surface ices, leaving dark, possibly carbonaceous, materials on the surface. They propose a cloudtop magnetic field of no more than 4×10^{-5} gauss.

Another proposal for a weak field will be put forward by Axford and Vasyliunas in next week's *Nature*. They argue that the magnetic field of Uranus originates in an

Jacques Oudin 1908–1985

WITH the death of Jacques Oudin on 15 October 1985, the international community of immunologists has lost one of its uncontested pioneers. From 1937, after completing his medical studies, Oudin spent his entire scientific career at the Pasteur Institute in Paris. From 1959 onwards he headed the laboratory of *Immuno-chimie Analytique* which was especially created for him. In 1979, the year after his retirement, a symposium at the Pasteur celebrated his exceptional career, highlighted by three major discoveries.

Some time after joining the Pasteur, while working in the laboratory of *Chimie Microbienne*, Jacques Oudin attempted to devise a one-step method to count the diverse antigen-antibody systems involved in the reaction of an appropriate antiserum and a complex mixture of antigens. He came up with the remarkable idea of using a gelled medium for the antigen-antibody precipitation reaction designing, in 1946, the method of immunochemical analysis. For a long time thereafter this was the only method for the enumeration and identification of antigens as well as the quantification of antigens and antibodies. Oudin formulated the quantitative laws and the general rules of the method, one of them — for which he had so often to fight — being that an antigen molecule gives only one precipitation zone with the corresponding antibodies.

It was while using antigen-antibody precipitates as immunizing material to elicit production of antibodies unispecific to the antigen present in the complex, that Jacques Oudin observed the first manifestation of what he defined in 1956 as allotypy of proteins, namely their property of exhibiting variations of antigenic specificities among individuals of an

animal species. Before the multichain structure of immunoglobulins was known, he showed the presence on the same rabbit immunoglobulin molecule of allotypic specificities governed by genes at two different and independent loci, which he had just identified. The discovery of allotypy opened the fertile field of immunogenetics and the fascinating area of immune-system manipulations.

Working along the same lines, Oudin found in 1963 that rabbit antibodies possessed additional antigenic specificities which seemed unique to antibodies to a given antigen in one individual or a group of individuals. He proposed the term *idiotypy* to designate them. This observation (made independently in 1963 for human antibodies by Kunkel, Mannik and Williams at the Rockefeller Institute) led to experiments which were crucial to our appreciation of both the potential repertoire of antibody diversity and the regulation of its expression. It also provided the basic material for Jerne's theory of the network of antibody variable-domain interactions and of the dynamic equilibrium of the immune system.

It is worth pointing out how far and wide the words coined by Oudin have spread: isotypy and isotype, allotypy and allotype, and particularly idiotypy and idio-type. But it would be presumptuous to try to capture with mere words the rigour of Jacques Oudin's scientific thinking, the peculiarity of his solitary creativeness, his talent in identifying important facts in unexpected findings that would have disconcerted less discerning minds, his intransigent empiricism, his perfectionism and the extreme care he took over all his writing. We have been privileged to benefit from such an exemplary mind.

Guy Bordenave

Earth-sized core, and they assume that the magnetic field from a body of this size must be about the same as the Earth's at the surface of the core. Thus, they conclude that the magnetic-field strength at cloudtop levels is about 10^{-2} gauss.

The proponents of a strong cloudtop magnetic field (≥ 0.6 gauss) for Uranus have become quiet apologists for the radio noise that has not been detected by Voyager: the radio emissions might be beamed away from the spacecraft; the aural electrons may be too low in energy (< 1 keV) to support the excitation of radio emissions; there may not be enough background thermal plasma for the auroral beams to become unstable to radiofrequency oscillations; the magnetic field of Uranus may be so strong that the radio emissions are above the sensitive range of the Voyager radio receivers; the magnetosphere may be rather quiescent so that energetic particles that lose their energy through synchrotron radiation are not

quickly replaced; and so forth.

Vociferous arguments in either side are lacking. The reason is plain — some of us are going to be embarrassed. Also, we have learned from other planetary encounters that nature is so clever at putting things together in completely unexpected ways that we now almost count on being astounded by what we find.

One cannot but wonder how, in a distant future, a community of, say, pulsar theorists will behave when an interstellar probe is about to fly by a pulsar for the first time. Rotating neutron star indeed! □

1. Clarke, J.T. *Astrophys. J.* **263**, L105 (1982).
2. Hill, T.W. & Dessler, A.J. *Science* **227**, 1466 (1985).
3. Desch, M.D. & Kaiser, M.L. *Nature* **310**, 755 (1984).
4. Curtis, S.A. & Ness, N.F. *Geophys. Res. Lett.* **12**, 855 (1985).

A.J. Dessler is Director of the Space Science Laboratory (ESOI), NASA Marshall Space Flight Center, Huntsville, Alabama 35812, USA.

Earth sciences

Perovskites and plate tectonics

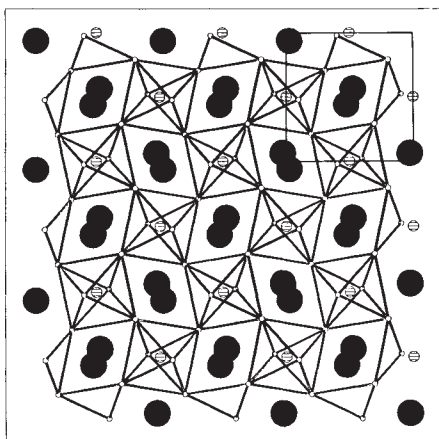
from Geoffrey D. Price

OUR ignorance of the dynamics of the Earth's mantle is such that it is still considered debatable whether the mantle convects as a whole, or in layers, divided one from another by a phase change or chemical boundary at the 670-km seismic discontinuity that separates the upper and lower mantle. Even though magnesium silicate perovskite probably makes up the vast majority of the lower mantle, and perhaps up to 40 per cent by volume of the entire Earth, little is known about its phase relations, its flow characteristics or its physical properties, all of which will critically influence the rheology and convective behaviour of the mantle. It is therefore essential to have a full knowledge of the nature of silicate perovskites if we are to understand the thermo-mechanical and chemical evolution of the Earth. The paper by Knittle *et al.* on page 214 of this issue attempts to provide the first data needed to begin such a task.

The lack of data on silicate perovskites is quite understandable, as it is not yet technically possible to measure the structural and physical properties of materials at the high pressures ($> 200,000$ atm) and temperatures ($> 2,000$ K) thought to characterize the lower mantle. To date, therefore, most of the theories concerning the behaviour of mantle-forming perovskite have been derived from the study of the behaviour of non-silicate perovskites. For example, O'Keeffe and Bovin (*Science* **206**, 599; 1979) as well as Poirier *et al.* (*Phys. Earth planet. Inter.* **32**, 273; 1983) have suggested that mag-

nesium silicate perovskite may behave like NaMgF_3 or KZnF_3 , which are solid electrolytes at high temperatures.

Until now, work on magnesium silicate perovskite itself has been confined to its synthesis, its structure determination and the study of its phase relations in the MgO-FeO-SiO_2 system. The study of the



A polyhedral projection of the orthorhombically distorted perovskite structure adopted by magnesium silicate perovskite under ambient conditions. Magnesium ions (large filled circles) sit in the distorted cages formed by the outlined corner-sharing SiO_6 octahedra.

thermal expansion of magnesium silicate perovskite by Knittle *et al.* is therefore to be welcomed as the first direct determination of any physical property of this mineral, and assumes particular significance when it is realized how difficult it is to use the diamond anvil cell to synthesize the quantity of perovskite that was needed for

the thermal expansion study. Such cells are notorious for producing impurities, because of the heterogeneous nature of the pressure and temperature distribution within the cell, but Knittle *et al.* have obviously overcome these problems to the extent that they have produced material which behaves in a systematic and reproducible fashion.

Because of the metastability of magnesium silicate perovskite under ambient pressure conditions, when studying its thermal expansion behaviour Knittle *et al.* were restricted to temperatures below 840 K. To address questions concerning mantle processes, they were therefore forced to extrapolate from their data to the pressures and temperatures found in the lower mantle. They are on relatively firm ground in doing this when discussing the variation in the physical properties of perovskites with changing environment. But, unfortunately, crystal chemical models are less robust to extrapolation than are equations of state, and consequently their conclusions concerning the possible changes in the structure of perovskite with increasing depth within the mantle are conjectural in the extreme. Silicate perovskites develop an orthorhombic distortion (see figure) under ambient conditions, but the nature of possible changes in the perovskite structure at high pressures and temperatures has excited considerable debate (for example, O'Keeffe, M. *et al. Phys. Chem. Minerals* **4**, 299; 1979), and Knittle *et al.* cannot really shed light on this aspect of perovskite behaviour.

From their analysis of the thermal expansion behaviour of silicate perovskite, Knittle *et al.* infer that the lower mantle must be enriched in iron or silica relative to the upper mantle, and reasonably conclude that the upper and lower mantle must convect separately, and that mantle-wide chemical mixing is impeded. The experimental difficulties involved in studying these high-density silicates are such, however, that the uncertainty in their measurements are large: indeed the results are compatible with perovskite thermal expansion coefficients that range from 2×10^{-5} to 6×10^{-5} K^{-1} . Although not discussed in detail, it is to be expected that such uncertainties are too large to enable the conclusions drawn by Knittle *et al.* to be considered definitive. Nevertheless, their approach must be pursued, as it will only be from such investigations that we will be able to determine the true nature of mantle behaviour and thereby understand the fundamental processes involved in plate tectonics and the evolution of the Earth. □

Geoffrey D. Price is a Royal Society 1983 University Research Fellow in the Department of Geological Sciences, University College London, London WC1E 6BT, UK.

Inositol phosphates

Profusion and confusion

from Bob Michell

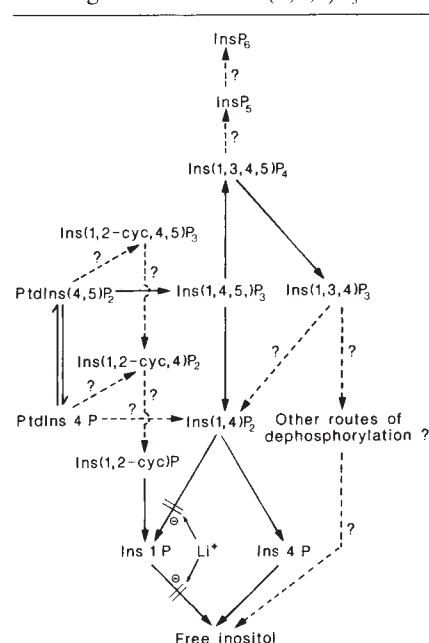
AFTER many years of argument, the past two years have brought agreement that receptor-stimulated hydrolysis of inositol phospholipids is a common mechanism for transmembrane signalling when cells respond to external stimuli such as hormones, neurotransmitters, antigens, light and some growth factors (reviewed in refs 1–3). But just as a slide depicting the simultaneous formation of two 'second messengers' within the cell by hydrolysis of phosphatidylinositol 4,5-bisphosphate ($\text{PtdIns}(4,5)\text{P}_2$) seems to have become a standard part of the travelling baggage of any scientist with an appreciable interest in cell responses to external stimuli, results from several laboratories suggest that this picture is greatly oversimplified. An up-to-date version of the formation and metabolism of inositol polyphosphate 'messengers' would need, unfortunately, to look something like that shown here.

The second messengers produced by the phosphoinositidase C (PIC)-catalysed hydrolysis of $\text{PtdIns}(4,5)\text{P}_2$ are *sn*-1,2-diacylglycerol and D-inositol-1,4,5-trisphosphate ($\text{Ins}(1,4,5)\text{P}_3$). The latter exerts its intracellular actions by releasing Ca^{2+} from an intracellular store, maybe in a compartment of the endoplasmic reticulum that is close to the plasma membrane², through interaction with an 'IP₃ receptor' that has now been detected by direct binding studies using liver microsomal membranes⁴. In solution at neutral pH, $\text{Ins}(1,4,5)\text{P}_3$ adopts the 'chair' configuration, with three phosphate groups lying around the periphery of the inositol ring⁵, and this may be the configuration recognized by its receptor. The major mechanism for terminating $\text{Ins}(1,4,5)\text{P}_3$ action is thought to be removal of the 5-phosphate by a family of specific phosphatases, present both in the plasma membranes and cytosol of stimulated cells. Although this picture of the formation, second messenger action and inactivation of the released inositol trisphosphate headgroup is almost certainly correct, it is now clear that several additional inositol polyphosphates, with as yet unknown functions, are produced in stimulated cells.

The first hint of this unexpected complexity came last year when Robin Irvine and Peter Downes discovered that much of the InsP_3 that accumulates in cholinergically stimulated parotid glands is the 1,3,4-isomer rather than the expected 1,4,5-isomer⁶; although the $\text{Ins}(1,4,5)\text{P}_3$ formation is immediate, that of the $\text{Ins}(1,3,4)\text{P}_3$ is delayed as is its decline on removal of the stimulus^{7,8}. The origin of $\text{Ins}(1,3,4)\text{P}_3$ was a mystery until the recent

evidence for its formation by dephosphorylation at the 5-position of D-inositol 1,3,4,5-tetrakisphosphate ($\text{Ins}(1,3,4,5)\text{P}_4$)¹¹, probably by what had been characterized as an $\text{Ins}(1,4,5)\text{P}_3$ -5-phosphatase^{9,10}.

$\text{Ins}(1,3,4,5)\text{P}_4$ is formed as rapidly as $\text{Ins}(1,4,5)\text{P}_3$ in various stimulated tissues and cells^{11,12} and at a recent meeting on inositol lipids in Germany*, Irvine and Michael Berridge presented evidence for a novel ATP-dependent kinase that converts $\text{Ins}(1,4,5)\text{P}_3$ to $\text{Ins}(1,3,4,5)\text{P}_4$. Because the absolute rates of synthesis and degradation of $\text{Ins}(1,4,5)\text{P}_3$ and of



Some known (→) and probable (---?) interconversions of inositol lipids and inositol phosphates in stimulated cells.

$\text{Ins}(1,3,4)\text{P}_3$ in stimulated cells appear to be of the same order, at least in the parotid gland⁷, the pathway for metabolism of $\text{Ins}(1,4,5)\text{P}_3$ through $\text{Ins}(1,3,4,5)\text{P}_4$ to $\text{Ins}(1,3,4)\text{P}_3$, and hence for its inactivation as a Ca^{2+} -mobilizing ligand, may be quantitatively as important as its dephosphorylation to $\text{Ins}(1,4)\text{P}_2$. Although it seems quite likely that $\text{Ins}(1,3,4)\text{P}_3$ and/or $\text{Ins}(1,3,4,5)\text{P}_4$ will turn out to have intracellular messenger functions of their own, this idea remains to be tested because so far the compounds have only been obtained in trace quantities.

Having found these unexpected inositol polyphosphates, the Cambridge groups looked further, and found that GH4 cells contain both an inositol pentaphosphate

(InsP_5) and an inositol hexaphosphate (InsP_6), although their concentrations do not change dramatically on stimulation of cells¹². There seem to be no previous reports of such highly phosphorylated inositol derivatives in mammalian cells, and their function and metabolism are unknown. However, InsP_6 , better known as phytate, is a major phosphate store in some seeds, and the 1,3,4,5,6-isomer of InsP_5 is an allosteric regulator of oxygen binding to haemoglobin in avian erythrocytes.

Almost 15 years ago, it was discovered that a major product of PIC hydrolysis of phosphatidylinositol in cytosolic and plasma membrane fractions from brain is inositol 1,2-cyclic monophosphate^{13–15} and that there is a widespread and specific phosphodiesterase that opens the 1,2-cyclic phosphate ring of this compound at the 2-position^{16,17}. As a result, inositol 1,2-cyclic phosphate was briefly considered as a possible intracellular messenger compound¹⁵. Philip Majerus and his colleagues in St Louis have now found that 1,2-cyclic derivatives can also be produced from $\text{PtdIns}(4,5)\text{P}_2$ and $\text{PtdIns}4\text{P}$ by two PICs isolated from sheep seminal vesicles, although in smaller quantities than the non-cyclic alternatives^{18,19}. D-inositol-1,2-cyc,4,5)trisphosphate ($\text{Ins}(1,2\text{-cyc},4,5)\text{P}_3$) is less potent than $\text{Ins}(1,4,5)\text{P}_3$ in mobilizing Ca^{2+} in permeabilized platelets but more so in depolarizing intact *Limulus* photoreceptors¹⁹. Although these observations show that $\text{Ins}(1,2\text{-cyc},4,5)\text{P}_3$ can mobilize Ca^{2+} , we do not yet know whether it is among the products of receptor-stimulated PIC activity in stimulated cells, because it would not survive the acidity of the media that are used for the efficient extraction of inositol trisphosphates from stimulated cells. If the 1,2-cyclic derivatives are formed in stimulated cells, it seems likely that they will be degraded, albeit more slowly, by the phosphatases that degrade $\text{Ins}(1,4,5)\text{P}_3$ and $\text{Ins}(1,4)\text{P}_2$, with the 1,2-cyclic ring being opened only after dephosphorylation to inositol 1,2-cyclic monophosphate²⁰. $\text{Ins}(1,2\text{-cyc},4,5)\text{P}_3$ may also be a substrate for $\text{Ins}(1,4,5)\text{P}_3$ 3-kinase, so giving rise to $\text{Ins}(1,2\text{-cyc},3,4,5)\text{P}_4$ and other cyclic phosphates derived therefrom.

A feature of the metabolism of receptor-generated inositol phosphates that has caused particular interest is its inhibition by Li^+ ions at extracellular concentrations similar to those that are effective in suppressing manic depressive symptoms. The dominant idea is that Li^+ inhibits $\text{Ins}1\text{p}$ phosphatase²¹ and thus may limit the supply of free inositol for inositol lipid synthesis (see ref. 2) and limit a cell's signalling capacity. Recent studies both *in vivo*²² and *in vitro*^{23,24} have, however, shown that $\text{Ins}(1,4)\text{P}_2$ can be degraded to inositol via inositol 4-phosphate as well as via inositol 1-phosphate. At least in liver,

*Inositol lipids and signal transduction, the 4th Ernst Klönk Foundation Conference, was held in Köln, FRG on 27–30 October 1985.

degradation via the 4-phosphate is Li^+ -insensitive, whereas both steps in degradation via inositol 1-phosphate are equally Li^+ -sensitive²⁴. The relative rates of $\text{Ins}(1,4)\text{P}_2$ degradation by these two routes are not known for any intact cells, but it seems probable that variations in this ratio will render signalling more sensitive to Li^+ in some cells than in others.

The point that emerges particularly forcefully from all of these recent findings is that the metabolism of inositol phosphates in stimulated cells is startlingly complex. It is clear that the very simple anion-exchange chromatographic methods that have been widely used for separating various inositol phosphate 'fractions' in recent years are quite unsuitable for analysing the complex mixtures of inositol phosphates present in stimulated cells. At present, the preferred method for more stringent analyses of the inositol metabolites present in stimulated cells is undoubtedly high-pressure liquid chromatography on anion-exchange resins. Three published methods^{7,19,25} each have some desirable features, but we can look forward to refinement of such methods and to the development of selective and sensitive assays for individual inositol phosphates. □

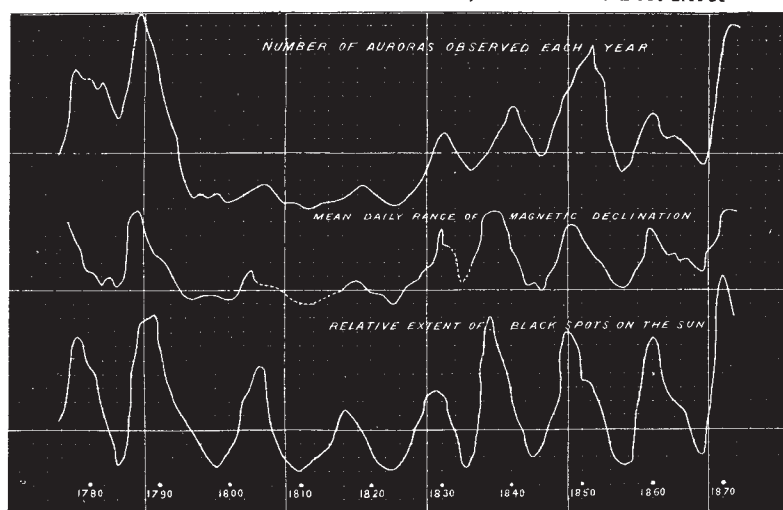
1. Nishizuka, Y. *Nature* **308**, 693 (1984).
2. Berridge, M.J. & Irvine, R.F. *Nature* **312**, 315 (1984).
3. Downes, C.P. & Michell, R.H. in *Molecular Mechanisms of Transmembrane Signalling* (eds Cohen, P. & Houslay, M.D.) 3 (Elsevier, Amsterdam, 1985).
4. Spat, A., Fabiato, A. & Rubin, R.P. *Biochem. J.* **233**, 929 (1986).
5. Lindon, J.C., Baker, D.J., Farrant, R.D. & Williams, J.M. *Biochem. J.* **233**, 275 (1986).
6. Irvine, R.F., Letcher, A.J., Lander, A.J. & Downes, C.P. *Biochem. J.* **223**, 237 (1984).
7. Irvine, R.F., Anggard, E.E., Letcher, A.J. & Downes, C.P. *Biochem. J.* **229**, 505 (1985).
8. Burgess, G.M., McKinney, J.S., Irvine, R.F. & Putney, J.W. *Biochem. J.* **232**, 237 (1985).
9. Downes, C.P., Mussat, M.C. & Michell, R.H. *Biochem. J.* **203**, 169 (1982).
10. Hawkins, P.T., Michell, R.H. & Kirk, C.J. *Biochem. J.* **218**, 785 (1984).
11. Batty, I.R., Nahorski, S.R. & Irvine, R.F. *Biochem. J.* **232**, 211 (1985).
12. Heslop, J.P., Irvine, R.F., Tashjian, A.T. & Berridge, M.J. *J. exp. Biol.* (in the press).
13. Jungawala, F.B., Freinkel, N. & Dawson, R.M.C. *Biochem. J.* **123**, 19 (1971).
14. Lapetina, E.G. & Michell, R.H. *Biochem. J.* **131**, 433 (1973).
15. Michell, R.H. & Lapetina, E.G. *Nature* **240**, 258 (1972).
16. Dawson, R.M.C. & Clarke, N. *Biochem. J.* **127**, 113 (1972).
17. Clarke, N. & Dawson, R.M.C. *Biochem. J.* **130**, 229 (1972).
18. Wilson, D.B., Bross, T.E., Sherman, W.R., Berger, R.A. & Majerus, P.W. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4013 (1985).
19. Wilson, D.B. *et al. J. biol. Chem.* (in the press).
20. Connolly, T.M., Wilson, D.B., Bross, T.E. & Majerus, P.W. *J. biol. Chem.* (in the press).
21. Hallcher, L.M. & Sherman, W.R. *J. biol. Chem.* **255**, 10896 (1980).
22. Sherman, W.R., Honchar, M.P. & Munsell, L.Y. in *Cyclitols and Phosphoinositides* (eds Bleasdale, J.E., Eichberg, J. & Hauser, G.) 49 (Humana, New York, 1985).
23. Storey, D.J., Shears, S.B., Kirk, C.J. & Michell, R.H. *Nature* **312**, 374 (1984).
24. Michell, R.H. *et al. in Hormones and Cell Regulation* Vol. 10 (eds Dumont, J. & Nunez, J.) (Elsevier, Amsterdam, in the press).
25. Binder, H., Weber, P.C. & Siess, W. *Analyt. Biochem.* **148**, 220 (1985).

Bob Michell is Reader in Biochemistry at the Department of Biochemistry, University of Birmingham, PO Box 363, Birmingham B15 2TT, UK.

100 years ago

from *Nature* **33** 252, 14 January 1886.

SOLAR SPOTS, MAGNETIC DECLINATION, AND AURORAL DISPLAYS.



Materials science

Limits to Coulomb's law

from Robert W. Cahn

MANY of the scientific laws that the readers of this journal absorbed in their youth are more than a century old and a surprising number go back more than two centuries — but which of them can still claim exact congruence with experimental facts? Some continue in daily use as sound approximations, for example, Boyle's, Hooke's and Ohm's laws, but lose their applicability in specific contexts, such as the behaviour of supercritical steam, 'potty putty' or semiconductor resistivity. Others, such as Newton's laws of motion, remain very close approximations to the behaviour of the real world until extreme conditions (for example, velocities) are attained. On page 203 of this issue, K. Kendall challenges another venerable law — Coulomb's law, formulated in 1773, which describes the frictional behaviour of a bed of powder.

According to Coulomb's law, the frictional force between a particle and a smooth plane pressed together by a normal force w is expressed as $f = k + \mu w$, where μ is the friction coefficient and k is a constant representing the intrinsic interaction — the van der Waals' attraction — between the mating surfaces. Coulomb's law has been used for a long time in engineering calculations, particularly in soil mechanics. Kendall now reports on experiments with wet sand confined between two silica plates and finds that the friction coefficient varies substantially, both with normal load and particle size.

The new findings imply that the van der Waals' interaction between, for example, two equal spheres depends on their diameters and on the force pressing them

together; this was first established by K.L. Johnson, K. Kendall and A.D. Roberts (*Proc. R. Soc. A* **324**, 301; 1971). If other factors are constant, the friction in a powder bed will rise rapidly as finer particles are used. In view of the widespread use of ultrafine powders in advanced ceramics, this finding has considerable practical importance. As with other named laws, Coulomb's law will presumably remain an adequate approximation under the normal circumstances of coarse powders and soil mechanics, but not under extreme conditions.

Coulomb's law is just one small feature of the varied, untidy but widely useful field of powder mechanics, a term not commonly used, but one that collectively covers dynamics, statics and thermodynamics. Consider some of the ways a powder can behave. When dry and in a still atmosphere, a powder has a 'flowability' that depends on particle mean size, size distribution, mean shape and shape distribution and on intrinsic frictional properties. An assembly of needles may stick obstinately, whereas dry uniformly-sized sand flows easily. The capillary effect of moisture is crucial: common salt needs a hygroscopic additive to keep it free-running; dry powder snow responds to pressure quite differently from wet snow, as any skier knows. If the proportion of water becomes too high, the result may be an avalanche or a quicksand.

A dry powder can be made fluid by blowing air or some other gas through it. Such a 'fluidized' bed has much lower friction than a normal powder bed and resembles a highly mobile liquid. The theory of

fluidized bed mechanics and heat transfer is important in chemical engineering, because such beds are commonly used for reacting solids with hot gases. When the gas pressure becomes too high, the agitation of the powder leads to erosion of the container. Such erosion has implications ranging from particle impact on the wind-screens of supersonic aircraft to the sand-blasting of castings or spark plugs. The abrasive wear of surfaces by powders has also become a large field of study.

The optimum feasible irregular packing of powders leads to complex statistical issues, which have been treated by J.D. Bernal and J.L. Finney (*Nature* **213**, 1079; 1967 and *Proc. R. Soc. A* **319**, 479; 1970). Their concept of dense random packing of equal-sized spheres, originally conceived as a model for monatomic liquids, has been extremely influential in guiding the modelling of metallic glasses. Surprisingly, a randomly packed assembly of mono-sized spheres can be within about five per cent of the density of a regular close-packed array. Dense random-packed arrays can be further densified by sintering, a technique that goes back to prehistoric practice in the baking of pottery. The stereology of solid-state sintering — the developing relation between packing fraction, porosity, permeability and specific surface area — has become a major field of scientific enquiry with relevance to metallurgy, ceramics and catalysis, and several novel experimental and conceptual tools have evolved with it. The mechanics of sintering — notably the crucial role of the curvatures of internal free surfaces and of intercrystalline boundaries — are now well understood. Through doping, subtle control of the motion of such boundaries has been achieved, making it possible to sinter to full density; the absence of scattering from pores has enabled the creation of translucent or transparent polycrystalline ceramics, now used in very efficient high-temperature lamps.

The ubiquity of powders in applied science indicates that Kendall's discovery of a major departure from a long-accepted powder law is of wider import than might at first appear. We still have to see how representative the circumstances of Kendall's experiments are. He used a wet load and the moisture fraction is not stated; presumably dry sand would behave quite differently. The plastic domain, as in hot isostatic pressing of metallic powder compacts, must entail frictional regimes quite different from those in the elastic regime treated by Kendall. Even so, his work has reopened a field that has been hemmed in by orthodoxy and neglected because ultrafine powders have only recently become commercially important. □

Robert W. Cahn is currently Fairchild Distinguished Scholar in the Keck Engineering Laboratories, California Institute of Technology, Pasadena, California 91125, USA.

Cosmic rays

Where do antiprotons come from?

from Andrew Buffington

It is usually believed that the antiprotons ($\bar{p}$) in cosmic rays result from collisions between high-energy protons (p) and ambient interstellar gas. Recent measurements of $\bar{p}$ cosmic-ray fluxes between 0.1 and 10 GeV are, however, inconsistent with such a secondary source, if the observed p fluxes are assumed to illuminate the $5-7 \text{ gm cm}^{-2}$ of interstellar material that has been deduced from studying heavier cosmic rays. Moreover, there is a kinematic cutoff for the reaction threshold of interstellar gas 'target-at-rest' interactions that predicts a dramatic drop in the $\bar{p}/p$ ratio below a few GeV, but the measurements show a practically constant ratio of 3×10^{-4} . The discrepancy factor at higher energy is about 5 but grows to about 100 close to 0.1 GeV. On page 205 of this issue C.D. Dermer and R. Ramaty¹ cite the recent $\bar{p}$ flux measurements and present a possible explanation for the reported similarity in $\bar{p}$ and p spectra.

Dermer and Ramaty suggest that the target material that is illuminated by high energy p to produce the $\bar{p}$ could be a relativistic plasma instead of cool interstellar gas. Such a plasma might be found in the vicinity of neutron stars or black holes. They then calculate that target particle motion is sufficient to remove most of the suppression of low energy $\bar{p}$ and show that production rate is significant for plasma temperatures above about $0.2 M_p c^2$. They discuss constraints put on the model by γ -ray measurements and suggest that many of the observed antiprotons were originally created as antineutrons, whose charge neutrality helped them to escape confinement in the ambient plasma magnetic field.

Many authors have already advanced a variety of explanations for the $\bar{p}$ anomaly with models that range from those invoking primary antimatter to those that adopt closed-galaxy cosmic-ray propagation or various types of shells around some of the cosmic-ray sources. Some of these are cited by Dermer and Ramaty, to which might be added an important calculation of secondary $\bar{p}$ production by Protheroe², models by Kiraly *et al.*³, Eichler⁴, and Lagage and Cesarsky⁵, and neutron oscillation as a possible separate source of low energy $\bar{p}$ ^{6,7}. I consider that the recent $\bar{p}$ flux measurements have stimulated significantly revised calculations of $\bar{p}$ production in thick targets. In these, the effect of further $\bar{p}$ collisions has been included, particularly because these can 'fill in' the kinematic suppression of the $\bar{p}/p$ ratio at low energy. Tau and Ng⁸ and Protheroe⁹ have shown that the observed $\bar{p}$ spectrum can be matched, provided the $\bar{p}$ were produced in

material with a thickness of about 30 gm cm^{-2} . In the case where all the $\bar{p}$ are secondary, the question becomes where might reside such a thickness for cosmic-ray protons to pass through and to what extent the heavier cosmic rays may also have passed through it. If the $\bar{p}$ are considered as secondary, our understanding of galactic cosmic rays stands at a major crossroads because $\bar{p}$ provide a unique tracer for the interactions of the primary p . (The spectrum of positrons, the other possible interaction-daughter candidate, are altered too much by photon interactions subsequent to their production to be useful.) The question is: should we adopt a different history for the plentiful cosmic-ray p , compared with that for the heavy cosmic rays?

We do not at present know what measurements could provide a definitive answer to this question but helium isotope spectra may be important as the next step. Because the helium-interaction mean free path is comparable to the 30 gm cm^{-2} of material indicated for parent-proton cosmic rays, considerable helium should survive the passage through even this thickness of material. As ^3He is expected to be rare in the cosmic-ray sources, the $^3\text{He}/^4\text{He}$ ratio shows whether helium cosmic rays have passed through a large amount of material, as some of the above models would predict. Recently, Jordan and Meyer¹⁰ reported the first measurement of $^3\text{He}/^4\text{He}$ near 6 GeV per nucleon, an energy high enough that little correction is needed for solar modulation. They report a ratio of 0.24, which requires about 15 gm cm^{-2} of interstellar material. On the other hand, measurements of this ratio below a few hundred MeV per nucleon are consistent with the $5-7 \text{ gm cm}^{-2}$ indicated by heavier cosmic rays, but in an energy range strongly affected by solar modulation. This discrepancy was the subject of lively discussion at a recent conference^{10,11}, so clearly Jordan and Meyer's result is not yet generally accepted.

These experiments are technically difficult, so that confirming and expanding experiments have been slow in coming, but this wave of new ideas about cosmic-ray origin and propagation needs new and better data to carry it along. It will be particularly important to have improved $\bar{p}$ data below 2 GeV, where the spectrum carries much information about re-interaction of the secondary $\bar{p}$; confirmation and extension of the high-energy helium isotope data is also needed. The group led by R. Golden at New Mexico State University has reconfigured their magnetic spec-

trometer, now ready for a new balloon flight, for $\bar{p}$ measurements to a lower energy range than they used previously; they have also made significant improvements in their trajectory measuring apparatus. New $\bar{p}$ instruments are also being planned by Goret¹² and by groups at Berkeley and Goddard^{13,14}. Although no high-energy $^3\text{He}/^4\text{He}$ experiments seem to be planned, the stimulating and exciting new results underscore their need. □

Andrew Buffington is at the Center for Astrophysics and Space Sciences, University of California, San Diego, California 92093, USA.

Marine reptiles

Lifestyle of plesiosaurs

from Michael A. Taylor

THE plesiosaurs, large extinct carnivorous marine reptiles of the Mesozoic period, have always excited dispute over how they swam¹ and reproduced. Did they swim like marine turtles, 'flying' with winglike, hydrofoil limbs, or did they row by moving their limbs back and forth (Fig. 1a, b)? And did they come on shore to lay eggs, or give birth in the water to live young? Recent research provides a new animal model that might answer these questions.

In 1975, Jane Ann Robinson argued that the plesiosaurs must have flown underwater, like turtles and penguins, because this movement was the most efficient use of their long hydrofoil-like limbs². If the plesiosaurs had rowed, they should have had webbed folding feet like those of ducks. She suggested that they swam by flapping both pairs of limbs up and down, producing lift force that propelled the animal forwards.

Seven years later Samuel Tarsitano and Jürgen Riess suggested a modification to Robinson's analysis³. Although agreeing that plesiosaurs must have used their limbs to fly underwater, they noted that the simplest — and apparently most effective — up-and-down action was anatomically impossible because the bones of the shoulder joints would block the appropriate operation of the forelimbs above the horizontal. Moreover, plesiosaurs do not seem to have had the large dorsal muscles needed to lift the wings powerfully upwards. Dino Frey and Jürgen Riess⁴ then suggested that the 'wings' had an active downstroke, and a comparatively passive upstroke as they were returned to the starting point. They suggested that the two pairs of limbs beat out of phase: one pair propelling the animal during its downstroke while the other was being brought back to the start of its own active stroke. This limb action would still be relatively inefficient compared with that of modern animals such as penguins,

1. Dermer, C.D. & Ramaty, R. *Nature* **319**, 205 (1986).
2. Protheroe, R.J. *Astrophys. J.* **251**, 387 (1981).
3. Kiraly, P. *et al. Nature* **293**, 120 (1981).
4. Eichler, D. *Nature* **295**, 391 (1982).
5. Lagage, P.O. & Cesarsky, C.J. *Astron. Astrophys.* **147**, 127 (1985).
6. Sawada, O. *et al. Astrophys. J.* **248**, 1162 (1981).
7. Sivaran, C. & Krishan, V. *Nature* **299**, 427 (1982).
8. Tau, L.C. & Ng, L.K. *J. Phys. G9* 227 (1983).
9. Jordan, S.P. & Meyer, P. *Phys. Rev. Lett.* **53**, 505 (1984).
10. Evanson, P. *et al. Proc. 19th Int. Cosmic Ray Conf.* **2**, 60 (1985).
11. Mewaldt, R.A. *Proc. 19th Int. Cosmic Ray Conf.* **2**, 69 (1985).
12. Goret, P. *Moriond Astrophys. Meeting, Cosmic Rays and Elementary Particles* (eds Audouze, J. & Tran Thanh Van, J.) 227 (1984).
13. Lloyd-Evans, J. *et al. Proc. 19th Int. Cosmic Ray Conf.* **3**, 254 (1985).
14. Solomon, M.H. *et al. Proc. 19th Int. Cosmic Ray Conf.* **3**, 258 (1985).

whose wings produce propulsive force throughout virtually the whole of the upstroke and downstroke.

Stephen Godfrey has now suggested a different living model for comparison⁵. The sealion swims by beating its forelimbs forcefully downwards and then relatively passively upwards, while always moving

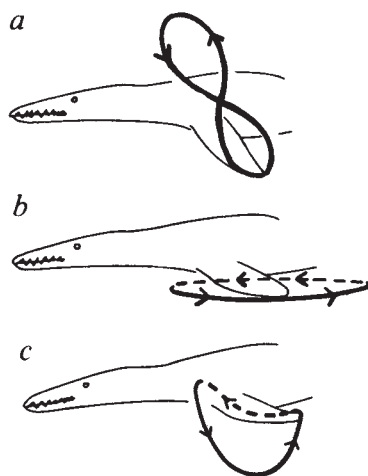


Fig 1. Possible mechanisms for swimming in plesiosaurs. The lines indicate the movement of the wingtip. *a*, Flying underwater like a penguin, with an up-and-down beat of the limbs; this movement would have been ideal for the limb shape but only the downstroke could have been really powerful. *b*, Rowing underwater like a duck, with an active power stroke backwards (bold line) then a recovery stroke forwards (dotted line). The plesiosaur's limb shape is particularly unsuited to such a movement. *c*, Swimming like a sealion, with a backwards power stroke and forwards recovery stroke, a movement combining elements of *a* and *b*.

them backwards. Once fully backwards, the limbs can then be feathered and brought forwards during the recovery stroke. Evidence points to the plesiosaurian shoulder and hip girdles having the massive muscles needed for such limb action (Fig. 1c). Ironically, sealion swim-

ming combines elements of underwater flight — up-and-down action — with rowing — backwards movement. The new model thus combines both the old theories.

If the plesiosaur had a swimming movement similar to that of the sealion, with a powerful downwards and backwards thrust, it is slightly easier to imagine how it pushed itself along a beach. However, the plesiosaurs with their rigid bodies could not have humped themselves along as sealions do but must have dragged themselves along, belly to sand, much as sea turtles now do. Whether they also laid eggs in the sand like turtles, or gave birth to live young in the water, as sea snakes do and ichthyosaurs are known to have done, is still unresolved.

The discovery of a pregnant female plesiosaur would settle the question, but none has yet been found. Perhaps we have not looked inside the bodies of enough plesiosaurs or perhaps the embryos decomposed before they could become fossilized. Two early reports of relevant fossils have recently been discounted. In the 1880s H. G. Seeley described an intriguing clutch of diminutive plesiosaur embryos from Whitby, Yorkshire, a discovery forgotten until R. A. Thulborn re-examined it⁶. The find turned out to be a forgery, an unusually shaped nodule (originally, he suggests, shrimp burrows) which had been 'improved' to strengthen resemblance to a clump of plesiosaur embryos, each several inches long. And a reported 'plesiosaur egg'⁷ is now thought to be an inorganic concretion⁸; even if it was an egg, it would probably be crocodilian, because crocodiles were much more common than plesiosaurs in the Upper Lias of Whitby⁹.

Tantalizingly, one clutch of eggs — but no young reptiles — was long ago reported from marine beach (or close inshore) deposits from the Middle Jurassic near Cirencester, England¹⁰. Again these were described as crocodilian rather than belonging to the much rarer plesiosaurs. What is needed, of course, is a clutch of eggs *in situ*, complete with tiny plesiosaurs; but such finds are exceedingly rare even for land animals such as dinosaurs¹⁰ and are even more unlikely to have been preserved on a wave-washed beach. □

1. Taylor, M.A. *Nature* **290**, 628 (1981).
2. Robinson, J.A. *N. Jb. Geol. Paläont. Abh.* **149**, 286 (1975); **153**, 86 (1977).
3. Tarsitano, S. & Riess, J. *N. Jb. Geol. Paläont. Abh.* **164**, 188 (1982).
4. Frey, D. & Riess, J. *N. Jb. Geol. Paläont. Abh.* **164**, 193 (1982).
5. Godfrey, S.J. *N. Jb. Geol. Paläont. Mh.* **for 1984**, 661 (1984).
6. Thulborn, R.A. *Palaeontology* **25**, 351 (1982).
7. Melmore, S. *Proc. Yorks. phil. Soc.* **for 1930**, 3 (1931).
8. Benton, M.J. & Taylor, M.A. *Proc. Yorks. geol. Soc.* **44**, 299 (1984).
9. Buckman, J. *Q.J. geol. Soc. Lond.* **16**, 107 (1860).
10. Horner, J.R. & Makela, R. *Nature* **282**, 296 (1982).

Michael A. Taylor works for the Area Museum Council for South-West England and is based at the City of Bristol Museum and Art Gallery, Queens Road, Bristol BS8 1RL, UK.

Neuroendocrinology

Molecular approach appraised

from Stafford Lightman

THE mechanisms controlling neuroendocrine regulation can be modulated by a variety of hormones, neurotransmitters and neuromodulators both *in vivo* and *in vitro*. As these systems have been well characterized by physiologists, they clearly offer an extremely attractive model in which to study the control of gene expression. The benefits and problems of the application of molecular genetics to neuroendocrine models were brought up to date at a recent meeting*.

Characterization of the precursors of hormones continues to be an important step in understanding the regulation of hormone production and in the discovery of new candidate hormones. A good example is the recent description of the precursor to gonadotropin-releasing hormone (GnRH) with its carboxy-terminal sequence GAP, which itself inhibits prolactin and stimulates gonadotropin secretion (Nikolics *et al.* *Nature* **316**, 511; 1985). Further studies (P. Seeburg, Genentech) confirm that GAP coexists with GnRH in hypothalamic cells, but also show that it is present in portal blood and even in the ovary and testis where GnRH has previously been demonstrated.

Except in frog skin, where it is found in abundance (Richter, D. *et al.* *EMBO J.* **3**, 617; 1984), the precursor to the tripeptide (Gln-His-Pro) releasing hormone, TRH, has been particularly difficult to identify. The problem has now been circumvented with antibodies to a synthetic peptide (Cys-Lys-Arg-Gln-His-Pro-Gly-Lys-Arg-Cys) presumed to represent a portion of the TRH precursor (R. Goodman, New England Medical Center, Boston). The antibodies were used to screen a DNA library from rat hypothalamus, which yielded a cDNA clone that encodes a protein containing five copies of the sequence Gln-His-Pro-Gly, flanked by pairs of basic amino acids. In support of the view that this protein is a TRH pre-

cursor, *in situ* hybridization with the cDNA was shown to detect neurones in those regions of the hypothalamus that are known to produce TRH. Thus, as with enkephalin, TRH has a precursor from which multiple copies of the biologically active peptide can be produced. Oddly, there is no homology between the frog-skin and rat hypothalamic precursors except for the TRH units themselves.

Defining the mechanisms of gene regulation is a major challenge to the molecular neuroendocrinologist. Quantitation of changes in pro-opiomelanocortin (POMC) mRNA in the anterior pituitary in response to corticotropin releasing factor and dexamethasone, or in the pars intermedia in response to bromocriptine, provide evidence of control at the level of gene transcription (J. Roberts, Columbia University, New York). The situation is more complex, however, in sex-hormone mediated changes in the arcuate nucleus of the hypothalamus. Treatment of a two-week ovariectomized adult rat with oestradiol results in a slow decline in the concentration of POMC mRNA in the arcuate nucleus and the expected corollary, a rapid decline in POMC transcription, which remains suppressed for days.

On the other hand, events controlling dopamine synthesis in the same system are not so clear. Oestradiol administration results within one hour in a fall of the transcription of the gene for tyrosine hydroxylase (TH) — the rate-limiting enzyme in catecholamine synthesis — with a gradual return to normal over 4 days, whereas the amount of TH mRNA begins to increase after 20 minutes and continues to rise for 4 days. This dissociation of the quantity of TH mRNA from the level of transcription of its gene implies that mRNA stabilization may be responsible for some of the quantitative changes in TH mRNA. Clearly, turnover of TH mRNA is the variable that we should like to measure, but the techniques are not available and, in particular, we have no idea how to study

the rate of specific mRNA degradation.

Similar problems apply to the interpretation of *in situ* hybridization studies, in which the hybridization of intracellular mRNA with specific DNA probes is used to locate sites of gene expression in tissue sections. This technique cannot provide any measure of RNA turnover and suffers from a lack of sensitivity which still limits its application to quantifying intracellular mRNAs. Nevertheless, it is finding many applications. For example, POMC mRNA can be detected in the pars intermedia of the rat pituitary, but in the anterior pituitary it is only found after adrenalectomy (J. Coghlan, University of Melbourne). Atrial natriuretic peptide mRNA could not be detected in the hypothalamus but has been found in the cardiac atria; renin mRNA has been detected for the first time in the arcuate nucleus. Vasopressin and oxytocin cDNA probes have now been used to measure mRNA responses to different physiological conditions (J. McCabe, Rockefeller University). Rats whose drinking water has been substituted with two per cent saline have an increased proportion of hypothalamic magnocellular cells heavily labelled by vasopressin cDNA, and magnocellular cells from 15-day pregnant rats show an increase both in the number of heavily labelled cells and in the number of cells with any autoradiographic grains. Further experiments are needed to show whether this represents two different groups of responding cells within the magnocellular paraventricular nucleus.

Novel techniques to measure gene expression and post-translational processing are exciting considerable attention. One method of incorporating mRNA into cells is by the use of a vaccinia virus (VV) vector (G. Thomas, Oregon Health Sciences University). The infection cycle of VV, unlike that of other DNA viruses, occurs solely in the cytoplasm of the host cell. Insertion of a cDNA downstream from an early VV promoter leads to transcription of the cDNA in the host cell within minutes of infection and to efficient production of protein from the mRNA. The potential of the technique is illustrated by the infection of different cell lines with recombinant VV containing the cDNA for human proenkephalin. Infected BSC-40 cells contain 2 main peaks of enkephalin immunoreactivity of which only the larger precursor is secreted into the medium. Infected AtT-20 cells, however, contain five peaks of immunoreactivity, with met-enkephalin as the major secreted form. This is a good example of tissue-specific processing. The considerable versatility of the VV system will doubtless provide much further information on this and other processing systems. □

*Neuroendocrine Molecular Biology, 13th annual meeting of the International Foundation for Biochemical Endocrinology, Edinburgh, 16–20 September, 1985.

Stamps of approval for Halley's comet



Stafford Lightman is Reader in Medicine at the Charing Cross and Westminster Medical School, London SW1P 2AP, UK.

Carbon cycle

Photosynthesis seen from above

from R.A. Warrick

ONE of the latest developments in remote sensing, reported by Tucker *et al.* on page 195 of this issue¹, is the capacity to estimate green-leaf biomass of plant canopies and, presumably, terrestrial photosynthesis and primary production, on a global scale with fine temporal and spatial resolution. This development has the potential to sharpen our understanding of the role of terrestrial biota in the carbon cycle and could allow the monitoring of important changes in global vegetation and productivity resulting from increasing atmospheric CO₂ and climatic changes.

As a result of fossil fuel burning, deforestation and other changes in land use, the atmospheric concentrations of CO₂ have risen from a pre-industrial (mid-nineteenth century) level of 275±10 to 343±1 parts per million by volume in 1984 (ref. 2). With a doubling of the pre-industrial CO₂ level the world could become considerably warmer, by 1.5–5.5°C according to climate modellers³ — the much-debated greenhouse effect. A major issue is when will this happen. The answer depends largely on future rates of energy consumption and CO₂ emissions, together with the sources and sinks of CO₂ and the fluxes between the various carbon reservoirs, which are the subjects of global carbon-cycle modelling. The terrestrial biota is a major component of such models.

In the presence of light, atmospheric CO₂ is assimilated by plants and transformed into carbohydrates — the process of primary production. Through respiration and the decay of organic matter, CO₂ is released back into the atmosphere. Because the rates of assimilation and release are dependent on solar radiation and temperature, and because the distribution of the terrestrial biota is lop-sided in favour of the northern hemisphere, the atmospheric CO₂ concentration has a strong seasonal cycle whose amplitude varies spatially. It has long been suspected that global-scale CO₂ variations are due to variations in primary production of the terrestrial biota, but the empirical demonstration and quantification of this has been hindered by insufficient data coverage, in terms of both time and space.

The remote-sensing techniques discussed by Tucker and his colleagues¹ take a large step towards remedying this problem. Their method involves advanced very high-resolution radiometer (AVHRR) sensors carried by the National Oceanic and Atmospheric Administration's polar orbiting meteorological satellites. The sensors collect radiance data daily, at 4 km spatial resolution with global coverage.

Daily collection is important, as it allows for selection of data for cloud-free days to minimize the effects of cloud cover. Tucker *et al.* have used these data to estimate the amount of intercepted photosynthetically active radiation (IPAR), the radiation in the spectral range that is strongly absorbed by leaves for driving photosynthesis. From this they estimate the green-leaf densities of plant canopies. If these factors provide an accurate measure of terrestrial photosynthesis, and if the annual CO₂ cycle is due mainly to photosynthesis variations, the satellite data should correlate well with atmospheric CO₂ variations.

The monthly, latitudinally-averaged values of a vegetation index derived from the data for the period 1982–84 have been compared with atmospheric CO₂ concentrations by Tucker *et al.* The patterns match well and show the expected inverse relationship: the higher the vegetation index, the lower the CO₂ concentration with a 1–2-month lag.

These remote-sensing techniques could potentially contribute to our understanding of the greenhouse problem in at least two ways. First, they could help in refining and validating global carbon-cycle models^{4,5}. There is considerable uncertainty about how well the terrestrial biota is represented by these models; few detailed validations have been performed². The global-wide data from the AVHRR sensors could provide an additional, independent data source against which CO₂ assimilation and release could be compared.

Second, they could be valuable for monitoring long-term effects of higher CO₂ concentrations and climatic change on the global vegetation. It is very uncertain whether and how the productivity of global ecosystems will be affected directly by higher CO₂ levels⁷ — the CO₂-fertilization effect. In carbon-cycle models this effect is often represented by a simple one-parameter equation in which the

assumed value of the biotic growth factor is estimated from results of greenhouse experiments. But scaling up the controlled experimental results from leaf or plant level to the real world of highly dynamic, interactive plant communities and ecosystems is fraught with difficulties⁸. The technique for monitoring changes in the pro-

ductivity of global ecosystems could provide the empirical data with which to test the assumption of CO₂ fertilization.

Change in species composition or areal distribution of ecosystems could also be caused directly by increasing CO₂ levels and by climatic change. Such vegetation changes have obvious feedbacks to carbon reservoirs and fluxes as well. Again, the satellite data could prove useful for detecting any such shifts.

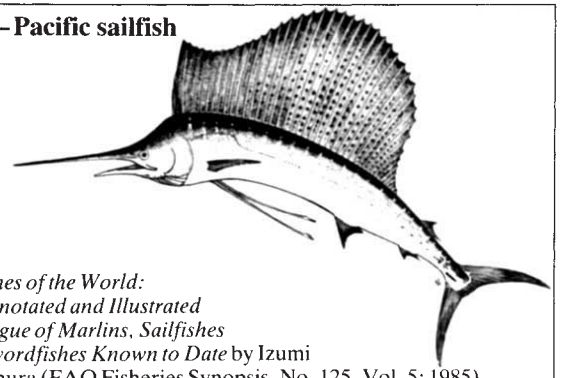
But the state-of-the-art in using the remote-sensing methods for relating reflected radiation, green-leaf biomass, primary production and atmospheric CO₂ levels is still immature. Tucker *et al.* note that unexplained variance and uncertainty exist at each link of this chain of relationships. How much confidence can we place in the approach? How much of an improvement over existing methods of measurement does it actually represent? Its biggest advantage is the global coverage, with fine resolution. But the sceptic must be convinced that we are not being sold sophisticated technology to address questions for which we largely know the answers.

Further careful comparisons of the satellite data with model results and ground-based data would help to establish the validity of the approach. Nevertheless, we have been given an enticing glimpse of a method with potentially far-reaching applications to the important global problems of the greenhouse effect and climatic change. □

1. Tucker, C.J., Fung, I.Y., Keeling, C.D. & Gammon, R.H. *Nature* **319**, 195 (1986).
2. Bolin, B. in *The Greenhouse Effect, Climatic Change and Ecosystems* (eds Bolin, B., Döös, B.R., Jäger, J. & Warrick, R.A.) (Wiley, Chichester, in the press).
3. Dickinson, R.E. in *The Greenhouse Effect, Climatic Change and Ecosystems* (eds Bolin, B., Döös, B.R., Jäger, J. & Warrick, R.A.) (Wiley, Chichester, in the press).
4. Björkström, A. in *The Global Carbon Cycle* (eds Bolin, B., Degens, E., Kempe, S. & Kettner, P.) 403 (Wiley, Chichester, 1979).
5. Emanuel, R.E., Killough, G.G., Post, W.M. & Shugart, H.H. *Ecology* **65**, 970 (1984).
6. Goudriaan, J. & Ketner, P. *Climatic Change* **6**, 167 (1984).
7. Wigley, T.M.L., Briffa, K.R. & Jones, P.D. *Nature* **312**, 102 (1984).
8. Shugart, H.H., Antonovsky, M.J., Jarvis, P.G. & Sandford, A.P. in *The Greenhouse Effect, Climatic Change and Ecosystems* (eds Bolin, B., Döös, B.R., Jäger, J. & Warrick, R.A.) (Wiley, Chichester, in the press).

R.A. Warrick is in the Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

Indo-Pacific sailfish



From
*Billfishes of the World:
An Annotated and Illustrated
Catalogue of Marlins, Sailfishes
and Swordfishes Known to Date* by Izumi
Nakamura (FAO Fisheries Synopsis, No. 125, Vol. 5; 1985).

Hydrogenases

From chemistry to legume growth

from R. Cammack and M.G. Yates

HYDROGENASES are enzymes used by bacteria and algae either to produce hydrogen or to consume it as a fuel. They are of interest not only as natural organometallic catalysts and because of the possibility of their application to biotechnological hydrogen production but also because hydrogenase genes may be used to increase agricultural productivity by improving the efficiency of nitrogen fixation. A recent symposium in Hungary* illustrated how the study of these fascinating enzymes has progressed.

It is now clear that there are two types of hydrogenase. The first contains iron, is extremely active and occurs in anaerobic bacteria such as *Clostridium*, which produce H_2 as an electron sink during their metabolism. The second type, one of the few enzymes known to require nickel, is more widely distributed and is used principally for H_2 uptake. It is less active in H_2 production than the Fe-hydrogenase, but from the biotechnological point of view it has the advantage of being much less sensitive to oxygen.

The structural genes of the Fe-hydrogenase of *Desulfovibrio vulgaris* have been sequenced (C. Veeger, Agricultural University, Wageningen). They encode a polypeptide of relative molecular mass (M_r) 46,000 and a hitherto undetected small subunit of M_r 13,500. The sequence indicates that two of the iron-sulphur clusters in this enzyme are of the conventional [4Fe-4S] kind and similar to those in bacterial ferredoxins; the third cluster is probably involved in catalysis.

The probable catalytic site of the Fe-hydrogenase from *Clostridium pasteurianum* has been scrutinized by Mössbauer spectroscopy and magnetic circular dichroism spectroscopy (L.E. Mortenson and M.W.W. Adams, Exxon Research, New Jersey; B.H. Huynh, Emory University, Atlanta; M.K. Johnson, Louisiana State University). It seems to be a unique type of four-iron cluster in which one iron atom is in a different environment but coupled to the other three.

The mechanism of action of the Ni-hydrogenases of *Desulfovibrio* species has been studied by electron-spin resonance and other spectroscopic techniques (J.J.G. Moura, Centro de Química Estrutural, Lisbon; J. LeGall, CEN Cadarache, France; R. C.). The nickel centre, which from EXAFS measurements seems to be surrounded by sulphur ligands,

can take up several oxidation levels including Ni(III) at an extraordinarily low redox potential. A hydride seems to be involved in the hydrogenase reaction mechanisms but its structure is not yet known.

What seems to be the commonest type of hydrogenase contains two subunits of M_s about 30,000 and 60,000. Important examples are the membrane-bound Ni-hydrogenases, which have been isolated from the hydrogen-oxidizing bacterium *Alcaligenes eutrophus* (K. Schneider and H.G. Schlegel, University of Göttingen), the aerobic nitrogen-fixing soil bacterium *Azotobacter vinelandii* and the symbiotic nitrogen-fixing organism *Rhizobium japonicum* (D.J. Arp, University of California, Riverside). These enzymes prove to be similar in structure and antigenic characteristics, as well as in regulatory factors (O_2 , H_2 and carbon substrates) and genetic determinants.

The genetics of hydrogenase has developed only recently. In *Escherichia coli* there are probably three different hydrogenases (D. Boxer, University of Dundee). Mutants in the structural genes have not been identified because a deficiency in one is masked by the presence of another, but several mutants in which the hydrogenase activity is affected have been isolated. Lee *et al.* have evidence of two classes of mutants, one of which contains *hydA* and *hydB* genes^{1,2}. The *hydB* gene is probably in a single operon with a gene encoding formate dehydrogenase and another for an electron-transport protein that couples formate dehydrogenase to hydrogenase. Of two newly discovered genes, *hydC* and *hydD* (L.F. Wu and M.A. Mandran-Berthelot, CNRS, Villeurbanne), the former may be involved in Ni processing.

The *hup* genes of *R. japonicum* Ni-hydrogenase have been isolated³. Plasmids containing these genes complement *Hup*⁻ mutants of *Azobacter chroococcum* and the arrangements of the *hup* genes in the genomes of these two species are similar⁴. The *hup* genes of *R. japonicum* are probably contiguous and comprise 15.5–21 kilobases divided into at least three transcriptional units⁵. The *hox* (hydrogen oxidation) genes of *Alcaligenes eutrophus* are plasmid-borne, encode both the membrane-bound and a soluble, NAD-linked hydrogenase and contain the structural and regulatory genes in a 50-kilobase sequence. However, an additional chromosomal function is necessary to express the Hox activity. This function may be part of a general regulatory

mechanism analogous to the *fnr* system of *E. coli* (B. Friedrich, Freie Universität, Berlin). The *hup* genes of *Rhizobium leguminosarum* are also plasmid-borne, those of *R. japonicum* may be, but those of *A. chroococcum* and one *Alcaligenes* strain are not; *hup* genes have also been isolated from the purple photosynthetic bacterium *Rhodospseudomonas capsulata* (A. Colbeau, A. Godfroy and P.M. Vignais, Centre d'Etudes Nucleaires, Grenoble).

Interest in hydrogenase genes is stimulated by the possibility that hydrogenase might improve the efficiency of biological nitrogen fixation by symbiotic rhizobia in leguminous plants, such as soybeans; many strains of rhizobia are naturally *Hup*⁻ or express low activity in legume nodules. The energy-intensive reaction of the enzyme nitrogenase produces hydrogen gas as an apparently inevitable by-product. If it builds up inside the soybean nodule, the hydrogen may inhibit the nitrogenase and it certainly represents a waste of energy resources. Hydrogenase, by recycling the hydrogen, could in principle alleviate these problems. But do *hup* genes produce bigger plants?

H.J. Evans and co-workers have investigated this question by comparing yields of soybean plants inoculated with *Hup*⁺ or *Hup*⁻ strains of *R. japonicum*. Both naturally occurring and genetically engineered *Hup*⁺ and *Hup*⁻ strains were used⁶. A range of responses was obtained with statistically significant higher yields in the *hup*⁺ series in most cases. Parallel work on free-living *A. chroococcum* (M.G. Y.) shows that its hydrogenase aids growth under selective conditions; these are carbon-limited, nitrogen-fixing continuous cultures at high dilution rates, or growth initiation in batch culture under air, together with low levels of carbon substrates.

The complexity of the *hup* operon appears to be comparable with that of the nitrogen-fixing *nif* operons. The fact that they have common regulatory factors (O_2 , fixed nitrogen and temperature) is probably coincidental, although *Hup*⁻ *Nif*⁻ mutants of *R. japonicum* are known⁷. Both the biochemistry and genetics of hydrogenase are important areas for continued research. □

1. Lee, J.H., Patel, P., Sankar, P. & Shanmugam, K.T. *J. Bact.* **162**, 344 (1985).
2. Sankar, P., Lee J.H. & Shanmugam, K.T. *J. Bact.* **162**, 353 (1985).
3. Lambert, G.R. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 3232 (1985).
4. Tibelius, K.H. & Yates M.G. in *Nitrogen Fixation Research Progress* (eds Evans, H.J., Bottomley, P.J. & Newton, W.E.) 536 (Nijhoff, Dordrecht, 1985).
5. Haugland, R.A. *et al. J. Bact.* **159**, 1006 (1984).
6. Evans, H.J. *et al. in Nitrogen Fixation and CO₂ Metabolism* (eds Ludden & Burris) 3 (Elsevier, Amsterdam, 1984).
7. Horn, S.S.M., Graham, L.A. & Maier, R.J. *J. Bact.* **161**, 882 (1985).

R. Cammack is in the Department of Biochemistry, King's College, (KQC), Campden Hill Road, London W8 7AH, UK; M.G. Yates is in the AFRC Unit of Nitrogen Fixation, University of Sussex, Brighton, BN1 9QJ, UK.

*Molecular Biology of Hydrogenases, Biological Research Centre, Hungarian Academy of Sciences, Szeged, 15–21 September, 1985.

The real cause of Ethiopia's problems

SIR—Constructive analysis and critical review by the scientific community of possible interventions to avoid future disasters in the drought-stricken African Sahel is the least that can be asked. John R. Krebs and Malcolm J. Coe¹ have certainly made a positive contribution to this process by assessing recent reports of the causes of the disaster, paying particular attention to the impressive argument of A.R.E. Sinclair and J.M. Fryxell² who proposed that the contesting causes are the weather and man. Krebs and Coe single out man and his recent efforts to develop sedentary agriculture as drawing the largest portion of the blame, but I suggest that much of their argument is not appropriate to the highland areas of Ethiopia so severely affected by the recent famine. Furthermore, the suggestion that a return to migratory lifestyle is the most viable option is unlikely to be considered feasible in those areas of Ethiopia.

Drought is not new to Ethiopia. However, Sinclair and Fryxell² argue that previous years of below-average rainfall did not result in the extreme mortality seen in the past 10 years. This is not the case. Ethiopia's chronicles and other historical sources reveal 23 major famines between 1540 and 1800, with those of 1540 and 1702 receiving particular comment as to their severity³. The famine of 1828–29 resulted in severe loss of life, and was complicated by an epidemic of cholera. The most devastating and well-described famine of 1888–92 reputedly killed one-third of the Ethiopian population⁴ and was complicated by a plague of locusts and an epidemic of rinderpest in cattle which almost destroyed the cattle population.

In many of the northern areas prone to cyclical drought, there is either no tradition of migration in search of pastures, water and cultivable soils or the migrations are distinct from the pastoral migrations more characteristic of other areas of Sahelian Africa. Sedentary agriculture in this area was established in the tropical highlands long before the advent of aid programmes which, along with colonization, have been proposed as major causes of the Sahelian famines^{1,2}. Ethiopia was never colonized, with the exception of a five-year occupation by Mussolini's forces. Extensive agricultural aid programmes are a relatively recent introduction and, judging from my three years' experience in Ethiopia a decade ago, have not had the impact, penetration and consequently destruction of the environment and the ecological balance that have occurred in many other parts of Sahelian Africa. Environmental destruction has of course occurred, as a result of high population density, overgrazing, denuding of

vegetation and soil erosion, but this has generally been self-inflicted and has occurred over a long time. The northern highland region of the country, also the most severely affected in the drought and famine of 1972–73, is relatively densely populated (more than double the average population density of sub-Saharan Africa), and much of it is characterized by plateaus at an altitude of 2,500–3,000m, gorges, mountain peaks and deep chasms. The peasant inhabitants are engaged in sedentary subsistence farming⁵, cultivating small plots of land, which as a result of centuries of feudal land tenure system and the traditional bequests of land, have become smaller with time. The farmers are hampered by the terrain and soil erosion occurs on an immense scale. Severe transport and communication problems associated with the rugged landscape compound their difficulties. They live, therefore, on the edge of survival.

The rainfall of the region is largely dependent on the movement of the Inter-tropical Convergence Zone (ITCZ) and its relation to pressure changes in the surrounding regions. Annual rainfall figures show very pronounced fluctuation from year to year, with variations of 56–140% of the mean figure reported during a 70-year period². Failure of the rains is therefore inseparable from the man-made components of this disaster. Furthermore, whereas overgrazing in highland Ethiopia is undeniably part of the man-made contribution to the cause, it is not the result of recent settlement near waterholes as suggested by Sinclair and Fryxell².

I submit therefore that in Ethiopia it is impossible to label the drought hypothesis as a misconception and to separate it from the alternative settlement-overgrazing hypothesis in discussion of the cause of the secular famines in that country. Furthermore, I think that broad generalizations of the Sahel which include Ethiopia should consider the geographical, historical, sociological and political factors which render it distinct from many other countries of the region.

Nevertheless, full support should be given to the recommendation of Sinclair and Fryxell that the improvement of the environment over several decades must be addressed by both developed and developing countries. The specific suggestions will require unprecedented collaboration between donor agencies and a close coordination between these agencies and the appropriate government departments in individual countries which, if to include such measures as cattle depopulation, will require a heavy hand to counteract traditional customs and practices relating to cattle keeping⁶. (For years veterinary services have contributed to the improvement of cattle production efficiency by reducing the prevalence of

many major infectious diseases but this has not been matched by corresponding improvement in livestock off-take and marketing.) But most important, the environmental improvement projects should be tailored to fit the individual countries, or in the case of Ethiopia, regions within the country, to avoid the undesirable effects of yet further inappropriate development aid such as that we are currently criticizing.

BRIAN D. PERRY

Virginia-Maryland Regional College
of Veterinary Medicine,
Blacksburg, Virginia 24061, USA

1. Krebs, J.R. & Coe, M.J. *Nature* 317, 13–14 (1985).
2. Sinclair, A.R.E. & Fryxell, J.M. *Can. J. Zool.* 63, 987–994 (1985).
3. Pankhurst, R. *J. Hist. Med.* 21, 95–124 (1966).
4. Pankhurst, R. *J. Hist. Med.* 21, 271–294 (1966).
5. Gryseels, G. & Anderson, F.M. *ILCA Res. Rep.* No. 4. (International Livestock Centre for Africa, Addis Ababa 1983).
6. Perry, B.D., Mwanaumo, B., Schels, H.F., Eicher, E. & Zaman, M.R. *Prev. Vet. Med.* 2, 633–653 (1984).

Major groove or minor groove?

SIR—After Matthews first proposed that cro protein has an α -helix that contacts the major groove of a right-handed B-DNA¹, it was soon realized² that the helix-major groove interaction might be a common feature by which many gene regulatory proteins recognize a particular sequence of DNA. Now we have the first crystallographic picture for this kind of binding in the studies on the phage 434 repressor-operator complex³. Although different in overall structure, endonuclease *EcoRI* has also been shown to have an α -helix contacting the major groove of its substrate⁴. I have suggested⁵ that stereochemically the major groove of B-DNA does provide the best place for a sequence-specific DNA-binding protein to make extensive contact with a 'recognition helix' as the main determinant of specificity⁶. The argument is as follows⁵. First, the diameter of an α -helix is at least 10.7 Å. Among common DNA structures, only the major groove of B-DNA is both broad and deep enough (about 11.7 Å and 8.5 Å) to accommodate an α -helix. Second, the major groove is much more specific than the minor groove for protein binding. This is because any type of mutation (A-T $\rightarrow$ T-A, A-T $\rightarrow$ G-C, A-T $\rightarrow$ C-G and C-G $\rightarrow$ G-C) changes the potential hydrogen bonding pattern much less in the minor groove compared to its effect on the pattern in the major groove. An extreme example is the A-T $\rightarrow$ T-A mutation, which keeps almost the same two H-donor positions in the minor groove but exerts a totally different hydrogen bonding pattern in the major groove. As for hydrophobic interaction, another important source of specificity, the main hydrophobic side chain (methyl

group of thymine) of DNA lies in the major groove as well.

Major groove contact is not universal, however. If the protein in question is not a sequence-specific one, the minor groove can also be its binding site, as reported by Klug and his colleagues in the histone core-DNA structure⁸. Also, proteins involved in transcription, reverse transcription and in the priming of DNA replication have a DNA-RNA hybrid as the binding partner. The hybrid molecule has been shown to assume A-form conformation⁹. Yet in A-conformation only the minor groove is open to proteins, while the major groove is too narrow and deep (2.7 Å in width and 13.5 Å in depth). Moreover, the two edges of the major groove are laced together by a continuous network of water molecules, making the major groove harder to access. It is then conceivable that the minor groove turns out to be the better binding site for those proteins. Since these DNA-binding proteins are not sequence-specific ones, the less specific minor groove is just good enough.

WANG JIA-HUAI

Biophysics Institute,
Academia Sinica, Beijing, China

1. Anderson, W.F., Ohlendorf, D.H., Takeda, Y. & Matthews, B.W. *Nature* **290**, 754-758 (1981).
2. Sauer, R.T., Yocum, R.R., Doolittle, R.F., Lewis, M. & Pabo, C.O. *Nature* **298**, 449-451 (1982).
3. Anderson, J.E., Ptashne, M. & Harrison, S.C. *Nature* **316**, 596-601 (1985).
4. Frederick, C.A., et al. *Nature* **309**, 327-331 (1984).
5. Wang, J.-H., in *The China-Japan bilateral symposium on biophysics* 142 (ed. Biophysics Society of China). (China Academic Publisher, 1985).
6. Wharton, R.P. & Ptashne, M. *Nature* **316**, 602-605 (1985).
7. Conner, B.N. et al. *Nature* **295**, 294-299 (1982).
8. Richmond, T.J. et al. *Nature* **311**, 532-537 (1984).
9. Wang, A.H.J. et al. *Nature* **299**, 601-604 (1982).

The facts on cystic fibrosis testing

SIR—The *News and Views* item¹ on a series of papers²⁻⁵ reporting DNA probes showing linkage to the cystic fibrosis (CF) gene observes "a start has at least been made on the development of the prenatal test for cystic fibrosis". It continues "the new data offer immediate prospects for the prenatal testing of cystic fibrosis". The reader would be entitled to conclude first that no prenatal test for CF currently exists and second that advances in molecular biology have radically altered this situation. Neither conclusion is correct.

Two years ago we reported that measurement of microvillar enzymes in amniotic fluid supernatant could be used as an *in utero* test for CF^{6,7}. Our data have been confirmed in many other laboratories and a special issue of the journal *Prenatal Diagnosis*⁸ was devoted to this subject. Although the test is not perfect, it has a very adequate detection rate of 95% at 18 weeks of pregnancy and a false positive rate of 5%. In the past two years more than 450 at-risk pregnancies have been prospectively screened by our procedures,

and the birth of over 100 children with CF prevented. With the possible exception of Tay-Sachs disease and β -thalassaemia, no other mendelian disorder has been as successfully tackled.

At first sight, it might seem that a restriction fragment length polymorphism (RFLP) tightly linked to the CF gene could improve prenatal testing by moving it from amniotic fluid samples in the second trimester of pregnancy to chorionic villus samples in the first trimester. However, this ignores a cardinal requirement of prenatal testing, namely that some method of confirmation of diagnosis on the abortus be available. This problem has been solved for the second-trimester fetus, first by the observation that most CF fetuses have the macroscopic appearance of meconium ileus⁹, and second by protein and enzyme analysis of meconium samples scraped from the fetal ileum⁹. These anatomical and biochemical changes are probably specific to the second trimester and observable only on an intact fetus. If they existed in the first-trimester fetus (which we doubt) they could not be assessed in materials removed by vacuum aspiration, the procedure for termination at this stage of pregnancy.

It is difficult to see how CF-linked RFLPs make reliable prenatal diagnosis 'a reality'. The critical question of the possible genetic heterogeneity of the disease has yet to be addressed. There are suggestions in two of the papers reviewed^{3,4} that the CF linkage relationship in the Amish may be different from that in other populations. Even if CF turns out to be a genetically homogeneous entity, much work needs to be done on producing highly informative flanking RFLPs, and on defining the extent of recombination expected with different types of multi-point analysis. But until the CF gene or its protein product is identified, the use of RFLPs on chorionic villus samples will lead to a form of prenatal testing without the benefits of confirmation of diagnosis. It must be emphasized that this is a radical departure from accepted norms of *in utero* testing.

We also have doubts about the value of RFLP-based heterozygote testing in CF. Here the need is for population-wide screening, so that heterozygous couples could be identified before the birth of the first affected child. Programmes of prenatal screening for Tay-Sachs disease and for β -thalassaemia have been effective, largely because heterozygosity tests have been simple, cheap and rapid. The current state of DNA technology is not yet suitable for this type of application.

There has been a tendency recently to exaggerate the immediate clinical benefits of advances in molecular biology. Careful assessment of the potential of RFLP-based diagnosis of genetic disease suggests

that its role is likely to be drastically limited by a number of factors such as genetic heterogeneity, the need to have information on whole families rather than individual patients, and expense. Although localization of a mutant gene on a particular chromosome will undoubtedly assist attempts at cloning that gene and identification of its protein product, it must be noted that this is yet to be achieved for any unsolved mendelian disorder, and may well prove more difficult than has been anticipated.

DAVID J.H. BROCK

VERONICA VAN HEYNINGEN

University of Edinburgh,
MRC Clinical and Population
Cytogenetics Unit,
Western General Hospital,
Edinburgh EH4 2XUf, UK

1. Newmark, P. *Nature* **318**, 309 (1985).
2. Knowlton, R.G. et al. *Nature* **318**, 380-382 (1985).
3. White, R. et al. *Nature* **318**, 382-384 (1985).
4. Wainwright, B.J. et al. *Nature* **318**, 384-385 (1985).
5. Tsui, L.-C. et al. *Science* **230**, 1054-1057 (1985).
6. Carbarns, N.J.B., Gosden, C. & Brock, D.J.H. *Lancet* **i**, 329-331 (1983).
7. Brock, D.J.H. *Lancet* **ii**, 941-943 (1983).
8. *Prenatal Diagnosis* **5**, 93-148 (1985).
9. Brock, D.J.H. et al. *Lancet* **i**, 1175-1178 (1985).

Long-running experiments

SIR—Your coverage of long-running experiments (*Nature* **318**, 310; 408 (1985)) brings to mind an experiment that although begun in this century will probably not be complete for millennia.

In 1918 Howard and Daniels¹ devised the simple but ingenious idea of sealing nitric oxide in glass tubes, some of which contained potential catalysts. Nitric oxide is thermodynamically unstable ($\Delta G^\circ_{f,298} = 86.6 \text{ kJ mol}^{-1}$) and any decomposition resulting from $2\text{NO} \rightarrow \text{N}_2 + \text{O}_2$ would be followed by $2\text{NO} + \text{O}_2 \rightarrow 2\text{NO}_2$ and the brown nitrogen dioxide produced would be easily detectable. When the tubes were examined 40 years later, decomposition was found neither in tubes containing pure NO nor in tubes containing NO plus catalyst.

The result was no surprise because an extrapolation of their higher temperature work in the presence of a calcium oxide catalyst indicated that 70,000 years would be required for 2% decomposition (the minimum they could detect) at room temperature. Howard and Daniels gave some of their tubes to the Smithsonian Institution, Washington DC, for future examination. In the case of the homogeneous gas phase decomposition of NO at room temperature, even if we assume a detection system sensitive to 100 molecules of NO_2 , it would still take $\sim 10^{12}$ years for any decomposition to be noticed.

SIDNEY TOBY

Physical Chemistry Laboratory,
South Parks Road, Oxford, UK

1. Howard, C.S. & Daniels, F. *J. phys. Chem.* **62**, 360-361 (1958).

What's in the name?

John A. Campbell

Artificial Intelligence: The Very Idea. By John Haugeland.
MIT Press: 1985. Pp. 289. \$14.95, £14.95.

AMONG scientific subjects, artificial intelligence has one unique distinction: it has been the victim of its own name. Because the name comes straight from everyday language, too many people have found futuristic messages in it merely through the associations that occur when "intelligence" and "artificial" are coupled together. The results have included science-fiction interpretations and unreasonable expectations of the subject, and confusion of what has been achieved with what has not. Quite rightly, some of the founding fathers have conceded, at least in private, that not enough care was taken 30 years ago to find a label free of loaded words.

Among specialists, there are at least three different definitions of what artificial intelligence involves. One is the study of human intelligence through the use of methods and metaphors depending essentially on computers and computation; this admits both a middle-of-the-road view, in which the definition means neither more nor less than what it says, and a man-in-the-black-box view, whose meaning is self-evident. A second definition is subtly different on paper but very different in practice: the computational study of tasks and processes which have previously been observed only in association with human intelligence. The third has arisen as a result of the subject's actual successes, and the observation that some concepts and techniques have recurred in many of the best pieces of work. By this modest operational definition, a project is an example of artificial intelligence if it relies mainly on those concepts or techniques. (A collection of this information, deliberately expressed in English rather than some high-level programming language, is given by P.H. Winston in the second edition of *Artificial Intelligence*, published by Addison-Wesley in 1984.)

The first definition is closest to the popular impression of what artificial intelligence is. It may even be a surprise to non-specialists to learn that the other two exist, the main reason being that books intended for a general audience and taking either of them as their starting-point have not been very visible in the past. Haugeland's book does not fill this gap. Instead, it joins the queue (or,

rather, the unruly crowd) attached to the first and most adventurous definition. It is, nevertheless, a welcome newcomer, because of its relative moderation and avoidance of science fiction.

As no exploration of the consequences of the definition can go far without touching on questions which were once regarded as lying exclusively within philosophical territory, it is an advantage that Haugeland's background is that of a philo-

But is there any future in Artificial Intelligence, Baron Frankenstein?



sopher. On previous evidence, such a vantage point has been no guarantee of successful writing on artificial intelligence — indeed, it seems rather to have guaranteed frequent encounters with the technical and even philosophical banana-skins that litter the field — but Haugeland arrives almost undamaged at the finishing line in this obstacle course.

His central concern is to explore the connection between intelligence and the idea of thought as a formal or at least rational manipulation of symbols. A first chapter on the evolution of philosophical views of "mind" is therefore followed by straightforward description of some of the symbol-manipulating systems, such as Turing machines, that have had the most influence on theoretical computer science and computability. There is also a simple account of models of computation, which makes the point that today's standard computer architecture, not well adapted to any of the definitions of artificial intelligence, is easy to replace with alternative designs that perform better in simulations. This is not news to specialists in advanced computing and to people familiar with the Japanese Fifth Generation plan, but it deserves repetition for a general audience.

As a basis for his substantive arguments, Haugeland's tour of the history of theoretical computer science is less satisfactory than his discussion of issues where he is surer of his philosophical ground. In the historical sections, for example, Alonzo Church is short-changed (by omission) on two counts and possibly three, with partial compensation for this Church-less picture only in the footnotes (where there is still confusion about the difference between Turing's "thesis" and Church's thesis, among other lesser things). Computer scientists may have the sense of seeing their subject through frosted glass here, but at least the author's heart is in the right place. And there is plenty of nourishment in the footnotes in general; collected conveniently at the end of the book, they can be sampled like a chariot of desserts.

The main value of the book to both specialists and non-specialists lies in the attention given to the two principal difficulties of current research in artificial intelligence. These are the frame problem, which is the problem of knowing how to link an item of information efficiently with all the other items that may be needed to establish the context in which its meaning may be determined properly, and the problem of how to assign and manipulate meanings (semantics) of expressions whose rules for structural (syntax) transformations are spelled out in formal symbol-manipulating systems. The problems themselves are well described, so that a general reader can appreciate the irritation that they cause to researchers, especially in speculation on how the human brain solves or avoids them. Haugeland weighs the balance of impressions justly: he is a pessimist rather than an optimist, but admits that, on the evidence, the basic questions of the achievability of "artificial intelligence" remain open.

While contexts and meanings may complicate work in artificial intelligence, there are both possible and desirable ways of reducing the complexity. On the "possible" level, problems of how to represent meaning may become much more tractable if the basic semantic operation used in a formal computational system, or even a brain, is nothing more than a two-way choice or distinction between alternatives or stimuli. And artificial-intelligence research can desire nothing better of its spectators than that they agree to replace the subject's label with one that calls up no wild associations. "AI" for "artificial intelligence", say? The very idea! □

John A. Campbell is Professor in the Department of Computer Science, University of Exeter, Exeter EX4 4PT, UK.

Two brains but of a single mind

John C. Marshall

The Duality of the Mind. By A.L. Wigan. Foreword by Joseph Bogen. *Joseph Simon, PO Box 4071, Malibu, California 90265, USA: 1985. Pp.346. \$35.*

THE question of how many angels can dance on the head of a pin has long since ceased to exercise our finest scholars. Yet if one wanted a contemporary analogue the following might suffice: how many minds may a man (or woman) manifest? "One", said René Descartes; "Two", said Arthur Wigan and Roger Sperry; "Forty-two", said Franz-Joseph Gall.

The answer "Two" was sanctified by the award of the 1981 Nobel Prize in Medicine and Physiology to Sperry for the work that he and his colleagues had undertaken on commissurotomy patients. Some 20 years earlier, the eminent Los Angeles neurosurgeon Joseph Bogen had succeeded in reducing the incidence of major convulsive seizures in a small group of severely epileptic patients by sectioning the main cortical fibre tracts (commissures) that connect the two hemispheres of the brain. Sperry's intensive testing of the cognitive functions of these patients' now-disconnected cortices sparked a veritable explosion of interest in the differential psychological capacities of the left and right cerebral hemispheres. On one account of the results, Bogen's operation did not *create* two minds within a single body but rather revealed that we are all dual creatures, deluded in our belief that each of us is an indivisible self.

Bogen's (or Bogens') own interest in these issues has extended beyond the investigation and interpretation of his "split-brain" patients to include the history of neurological speculation about cerebral duality. One fruit of his delving into the past is this splendidly handsome reprint of Arthur Wigan's Victorian epic of 1844, the scope and character of which is best shown by its full title: *A New View of Insanity: The Duality of the Mind Proved by the Structure, Functions, and Diseases of the Brain and by the Phenomena of Mental Derangement, and Shown to be Essential to Moral Responsibility*. Indeed, Wigan's version of the "two-brain" theory is sufficiently close to some modern accounts for Bogen to name his own position "neowiganism".

Who might want to read the book today? To begin with, anyone who is concerned with the relationship between the neurosciences and society will be fascinated to see Wigan struggling to resolve the tension between biological determinism, educability and moral choice. Wigan's

phraseology, and the certainty of his Victorian belief in scientific and ethical progress, may seem quaint to our more cynical ears:

The only limit to our researches on the nature of the mind, will ultimately be the boundary fixed by the Almighty to the powers of the human intellect — a point from which we are yet immeasurably distant. When we shall have cultivated all the faculties which He has bestowed upon us, to their full extent and perfection, then indeed will come the Millennium — an issue towards which we are steadily and rapidly advancing. . . .

yet his discussion of the potential conflicts between humanistic and scientific approaches to psychiatry and neurology is a good deal more honest and open than many current positions.

Second, the book *should* be read by all students of neuropsychology. Wigan is a pre-modern in that his work pre-dates the discovery of complementary hemispheric specialization that became (and remains) the central dogma of human neuro-

psychology. Thus the notion that the left hemisphere is primarily implicated in language functions and the right in visuo-spatial functions is quite alien to Wigan's time. For Wigan, the hemispheres are full duplicates with respect to their intrinsic cognitive capacities; it is solely "slight inequalities" of "form, energy, and function" that suffice "to produce all the varieties of character which are to be found in the world". Accordingly, many of the case reports that Wigan describes are directed towards showing that "One cerebrum may be destroyed, yet the mind remain entire". The contemporary student would benefit greatly from the exercise of contemplating whether Wigan's conclusion derives merely from inadequate testing of non-verbal skills or rather contains an important truth that has been buried under our own obsession with complementary specialization. □

John C. Marshall is in the Neuropsychology Unit, part of the Neuroscience Group at the Radcliffe Infirmary, Oxford OX2 6HE, UK.

A southern man

Derek Fordham

Shackleton. By Roland Huntford. *Hodder & Stoughton/Atheneum: 1985. Pp.774. £19.95, \$29.95.*

"MEN wanted for hazardous journey. Small wages, bitter cold, long months of complete darkness, constant danger. Safe return doubtful. Honour and recognition in case of success." Thus Sir Ernest Shackleton chose to launch his 1915 Imperial Trans-Antarctic Expedition and was overwhelmed by thousands who wanted to join.

Years earlier, in 1900, Shackleton had been in the Antarctic for the first time and Roland Huntford spares no one in setting out the dreadful deficiencies and incompetence of that *Discovery* expedition under the leadership of Commander Robert Falcon Scott. During the expedition Shackleton came close to death and subsequently suffered the ignominy of being invalidated home on Scott's orders. Scott's action was undeniably motivated by his dislike of Shackleton, who had shown himself to be a natural leader, in which respect Scott rightly doubted his own abilities.

In a spirit of defiant rivalry, heavily in debt and with a suspected weak heart, Shackleton set about organizing his own expedition, which sailed south in 1907. Shackleton and his team, poorly equipped and ill-prepared, penetrated to within 97 miles of the Pole before he decided to turn back, the hardest decision he ever had to make, but inescapable if he was to bring his team back alive. In so doing he laid

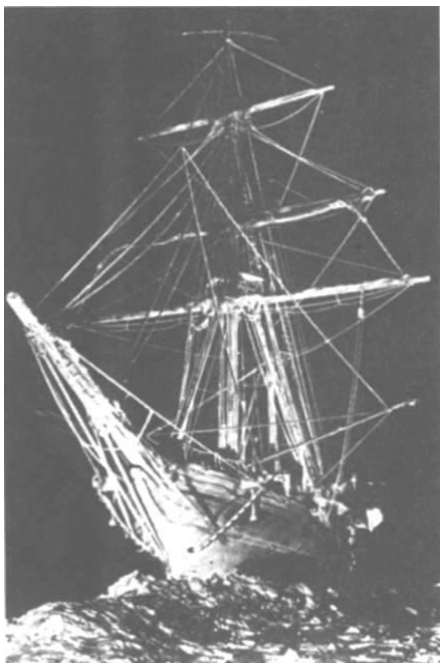
bare the geography of the polar plateau and opened the route to the Pole for Scott and Amundsen to follow in 1911. His hero's return was marred by financial difficulties and the scandal of an errant brother, suspected of stealing the Irish crown jewels.

His resources in the face of failure were taxed again in October 1915 at the start of The Imperial Trans-Antarctic Expedition, when his ship, *Endurance*, having drifted helpless for 1,500 miles trapped in the pack-ice, was crushed and sunk in the Weddell Sea. For 600 miles Shackleton led his men over shifting, breaking ice to the floe edge and then across the storm-swept open sea to the desolation of Elephant Island. Having established his men in a primitive camp he set off a few days later with five companions on a further 700 mile open boat journey across the Southern Ocean to South Georgia. A final desperate traverse of a range of unmaped icy mountains led to the whaling station from which the rescue of his expedition was mounted. Shackleton the survivor had once again brought his men back from the prospect of certain death.

But a world on the brink of war failed to recognize the staggering accomplishment of this impetuous man and he had to struggle to raise further funds for his last expedition, during which the misfortunes encountered on his earlier Antarctic forays took their toll. He died as his ship reached South Georgia, ironically the scene of his return from the dead in 1916.

This is Roland Huntford's second polar book — the first, published in 1979, having left the schoolboy image of Captain Robert Falcon Scott somewhat mangled and the subject of much controversy — and deals with the very different but no

less intriguing figure of a man who mixed Irish dreaming and poetry with the power to inspire those who worked with and for him in a manner previously unknown in polar exploration. The differences between Shackleton and Scott are developed and the author sees the comparison between the upright, unsure naval officer and the breezy, extrovert Irish poet as illustrating how, in an age before technology came to dominate polar exploration, Shackleton's great achievement was to break by sheer power of leadership and commitment to his objectives the



Ice-bound: Endurance trapped by pack-ice in the Weddell Sea.

mediocrity that had until then laid its dead hand over much British polar activity. It was not without reason his men called him simply, "Boss".

This is an impressive work and will surely stimulate new interest in a man whose achievements in the field of polar exploration have down the years been largely overshadowed by the tragic nature of his former leader's death.

On the crossing of South Georgia, which ended the journey for which above all else the world should remember Shackleton, he and his two companions felt they were accompanied by an extra person. This is a not uncommon experience for explorers under stress, and Shackleton's account of this phenomenon inspired a haunting reference in T.S. Eliot's poem, *The Wasteland*, "Who is the third who walks always behind you?".

Perhaps for "The Boss" it was the embodiment of the conventional hero, which his enigmatic spirit never quite embraced. □

Derek Fordham, 66 Ashburnham Grove, Greenwich, London SE10 8UJ, UK, is Secretary of the Arctic Club.

States of alteration

C. Ladd Prosser

Changes in Eukaryotic Gene Expression in Response to Environmental Stress. Edited by B.G. Atkinson and D.B. Walden. Academic: 1985. Pp.378. \$71.50, £60.

The Adaptive Role of Lipids in Biological Systems. By Neil Hadley. Wiley: 1985. Pp.319. \$39.95, £41.40.

ATKINSON and Walden are concerned with the control of protein synthesis and, as the basis for discussion, the contributors to their book have focused on proteins made in response to stress. It was observed 30 years ago that brief exposure to heat brings about puffs on salivary chromosomes of *Drosophila* and leads to the synthesis of several proteins in large quantity — the so-called heat-shock proteins. In *Drosophila*, seven proteins of different molecular weights have been identified and several have now been cloned. Although most information on heat-shock proteins is available for *Drosophila*, and this is reflected in the book, other chapters describe heat-shock proteins in a slime mould, sea urchin tissues, amphibians, birds and mammals, and in plants.

The heat-shock proteins in different organisms are of various sizes but the most prevalent one is of 70 K molecular weight. They are highly conserved and can be induced by heat exposure of isolated cells as well as of intact organisms. Control has implicated both translation and transcription but evidence from inhibitors emphasizes the increase in transcription of specific mRNAs as most critical. It appears that heat exposure turns on certain genes and one effect of the synthesis of heat-shock proteins is to repress action of mRNAs for other proteins. The chapter by Lindquist and Didomercio on gene expression in *Drosophila* uses synthesis of the mRNA for the 70 K protein as a model for gene regulation in general. The *in vitro* induction of heat-shock proteins in avian red blood cells is covered in chapter by Atkinson and Dean. In mammals, a fever of only 2°C is sufficient to induce heat-shock proteins in the brain.

Synthesis of these proteins is stimulated by other stresses, particularly toxins and drugs, hence the term stress proteins may be more appropriate than heat-shock proteins. Less is known of their synthesis in plants than in other organisms, but maize seedlings and soybean embryos do produce them. There seems no doubt that synthesis of heat-shock proteins is associated with increased resistance to heat stress. Although much is known of the sequence of specific gene activation and protein synthesis, the mechanisms by which heat-shock proteins protect against the lethal action of heat remain mostly

speculative. But Atkinson and Walden have succeeded well in integrating the groundwork essential for a final answer to the question of mechanism within the seventeen chapters of the book.

Adaptation in a different system is the topic of the volume by Hadley. He has considered the adaptive function of lipids in a number of biological roles. The book could serve well for one part of a course in comparative physiology. The treatment of biochemistry of lipids is adequate for physiology students but probably not sufficient for lipid chemists. Lipids, predominantly phospholipids, constitute 25–75% of cell membranes by weight. The lipid composition of plasma and organelle membranes can be altered by nutrition and by temperature, but the way in which membrane lipids modify the functioning of membrane proteins remains an unsolved problem.

A useful chapter on surface waxes precedes a detailed discussion of the role of waxes in the water relations of terrestrial plants and animals. Their protective role against evaporative loss of water is made clear, but the significance of the diversity in composition of surface waxes — hydrocarbons, ketones, esters, alcohols, sterole and triterpenes — is not explained.

Lipids, especially triglycerols, are stored as fuel that can be accumulated at certain times, for example at the pre-migration stage in insects and birds. Several roles for fat tissue are further described. Brown fat is a ready source of heat, present in most mammalian neonates and hibernants. Because fat has low thermal conductivity it is an insulator, especially in aquatic and cold-dwelling mammals. Some marine plants and animals are neutrally buoyant because of low-density fats. Buoyancy is implemented by an increase in total body lipids and by deposit of lipids of low density. A use of lipids not often considered is in chemical communication, especially in pheromones; some esters of plants are insect repellents. A final chapter deals with medical aspects — pulmonary surfactants, bile salts, fat absorption.

The overall impression to be gained from this book is of the extreme diversity of structure and function of biolipids. Clearly, this is a field of research in which comparative biochemists and ecophysiologists can profitably cooperate. □

C. Ladd Prosser is Emeritus Professor of Physiology in the Department of Physiology and Biophysics, University of Illinois at Urbana-Champaign, 524 Burrill Hall, 407 South Goodwin Avenue, Urbana, Illinois 61801, USA.

• *Ancestors: The Hard Evidence*, the proceedings of a symposium held in April 1984, has been published by Alan R. Liss/Wiley. The meeting was attended by leading workers in palaeoanthropology and was reported in *Nature* 208, 309 (1984). Editor of the book is Eric Delson, price is \$49.50, £38.

Talking of muscles and motion

R.M. Simmons

Design and Performance of Muscular Systems. Edited by C.R. Taylor, E. Weibel and L. Bolis. *The Company of Biologists Limited*: 1985. Pp.412. £25, \$60. Available from the *Biochemical Society Book Depot*, PO Box 32, Commerce Way, Colchester CO2 2HP, UK.

ONE of the main themes of this book, which is based on papers presented at the Seventh International Conference on Comparative Physiology in June 1984, concerns problems of gait during locomotion, such as why animals adopt narrow ranges of speed for particular gaits and why they change from one gait to another (Cavagna, Hedlund, McMahon and Taylor). Some of the illustrations are drawn from unusual human gaits such as hopping, Groucho running (that is, with very bent knees like Groucho Marx) and running on very compliant surfaces.

One idea, now 10 years old, is that an animal can be considered during locomotion as a multi-jointed mass-spring system with a characteristic resonant frequency, and thus a preferred stride rate. As speed increases, say in walking, the ride becomes uncomfortable and less economical because of the stiffness of the spring arrangement and running is more comfortable as the ride is softer and it is less costly in energy consumption. Maximum speed might be determined by maximum stress and this varies with body size much as expected (Alexander). Under the mass-spring model, unusual gaits extend the range of observations.

In locomotion on a flat surface the net external work done is small, but groups of muscles alternately shorten and are stretched thus performing positive and negative work respectively. The tuned-spring model implies that energy imparted to muscles by stretch is recovered when they shorten; indeed, this was first demonstrated in Margaria's laboratory 20 years ago. The positive work phase can thus appear surprisingly efficient, up to 70% in large animals, though much lower in smaller animals for reasons that are not entirely clear. More details of the process of energy storage come from experiments on isolated muscles (by Cavagna and his colleagues), showing increased energy output when an actively contracting muscle is first stretched before being allowed to shorten, some of the power coming from elastic energy stored in the tendons and some apparently stored in the cross-bridges.

Another theme in the book concerns the properties and determinants of the

two main muscle fibre types in mammals, the fast-twitch fibres whose mechanical properties are fast and whose metabolism is mainly glycolytic, and the slow-twitch fibres which are slower and are mainly oxidative. The current state of knowledge about structural protein isoforms is given by Perry and Whalen, but most of the new findings about the proteins in mammalian muscle are concerned with mitochondrial enzymes: in separate studies Pette and Edgerton agree that the heterogeneity found in a given fibre type implies a different mode of specification than for the structural proteins, but disagree about whether fibres innervated by branches of the same nerve show heterogeneity. Kushmerick presents his now completed study of the energetics of isometric contractions in two muscles from the mouse, one predominantly slow and the other fast.

A third theme concerns the malleability of the system. Muscles predominantly of one fibre type can be converted into the other by cross-innervation or by changing the pattern of muscle activation artificially: continuous low-frequency stimulation yields slow-twitch characteristics, whereas occasional bursts of stimulation make for fast-twitch fibres. The major determinant of muscle speed is the myosin isoform, while the duration and rapidity of response to a stimulus is probably determined by the membrane systems.

From painstaking stereological measurements it appears that the membrane systems change before the structural proteins (Eisenberg): this can induce an abnormally high mitochondrial content in a fast to slow transformation, the muscle's response to the energetic demands caused by more frequent stimulation and longer response times while fast-cycling myosin is still present. A similar type of mismatch arises during early post-natal development in the rat if a nerve is crushed and then allowed to re-innervate a muscle (Vrobova). The pattern of motor-neurone activity is changing during this period, so that by the time re-innervation

is complete the immature muscle encounters a mature rate of stimulation and partial atrophy results. In the adult, complete recovery results, the reverse of what might be expected intuitively of variation in plasticity with age.

The sequence of appearance of myosin isoforms during development now seems well established (Whalen). Development from embryonic through neonatal to fast myosin occurs whether or not the muscle is innervated, but induction of slow myosin is nerve-driven. However, thyroid hormone is needed for the transition from neonatal to adult fast myosin. A recent report (discussed here both by Howald and Goldspink) describes changes in myosin isoform from slow to fast as a consequence of training in the rat, but changes in man seem to be less pronounced. Successful athletes in endurance events, such as long-distance runners, have a preponderance of slow twitch fibres and sprinters have more fast twitch fibres. Training in man can induce selective fibre type hypertrophy and an increase in mitochondrial content (Gollnick, Hoppeler), but it is still not clear whether an athlete's fibre typology selects the best event for him or the athlete selects the typology for his chosen event by training.

Other aspects of muscle performance such as whole body response and the control of movement are also covered by appropriate experts. There are in all 34 articles, mostly by different authors, and although the introductory chapters (by H. Huxley, Peachey and Wilkie as well as those by Perry and Whalen already mentioned) are good, I think the novice might find it hard to follow the threads of some of the arguments. For the professional and others with some background in the fundamental properties of muscle, however, this book offers some valuable insights into the design and performance of muscular systems. □

R.M. Simmons is Head of the Department of Biophysics, Cell and Molecular Biology at King's College London (KQC), 26-29 Drury Lane, London, WC2B 5RL, UK.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Running out of time: Deinonychus, a medium-sized carnivorous dinosaur from the Early Cretaceous Cloverly Formation of Montana, reconstructed by Robert T. Bakker.

The illustration is reproduced from Bones for Barnum Brown: Adventures of a Dinosaur Hunter

by Roland T. Bird, edited by V. Theodore Schreiber. Publisher is Texas Christian University Press (and distributor Texas A & M University Press), price is hbk \$29.95, pbk \$14.95.

Edwin Hubble: legal eagle

from N. Hetherington

Edwin Hubble skilfully employed trial tactics and techniques to attain favourable verdicts from the court of science.

THERE is more to the advance of science than new observations and new theories. Ultimately, people must be persuaded.

Galileo, a master polemicist, appealed skilfully and effectively to a public forum when science, yet to be differentiated from philosophy, was still accessible to most educated people. An explosion of knowledge and specialization has since narrowed the body from which competent jurors can be drawn for the trial of a scientific idea. Yet, whether before the public or before a jury of their peers, scientists are still concerned with persuasive presentation.

The twentieth-century astronomer Edwin Hubble well appreciated the advocative aspect of science and was a skilled practitioner. Reading Roman law at Oxford and practising as an attorney in the United States were good preparation for a subsequent career in science. Trial tactics and techniques gained for Hubble favourable verdicts more quickly than might otherwise have been attained.

Independent witnesses

Hubble's bold extrapolations from near to distant galaxies presumed uniformity within the universe, a uniformity somewhat supported by Hubble's own studies at the Mount Wilson Observatory of the distribution of absolute magnitudes among isolated nebulae. Harlow Shapley at Harvard, however, had reasonable doubts. Hubble, attempting to persuade doubters with independent corroboration, emphasized, perhaps even exaggerated, the independence of Nick Mayall's study at the Lick Observatory.

Hubble had supported Mayall's dissertation proposal, and Mayall's analysis of nebular counts turned out to agree reasonably well with Hubble's own data¹. When it came to publishing Mayall's results, though, Hubble contrived to have the study appear as a completely independent body of work. Thus it might more strongly persuade astronomers in the eastern United States, including Shapley. Hubble wrote to Mayall²:

your investigation should be considered as absolutely independent of our work here. Actually it would have greater influence on the eastern gang if there were a note of challenge.

Thus was Mayall's evidence presented in court, to Hubble's satisfaction³:

I am glad to hear that the thesis will be in print soon and thoroughly agree with the committee that it should be published

without reference to my paper. Actually I have a personal interest in this decision for I am eager to have as much independent investigation as possible so that the accumulating evidence will eventually convince the eastern group of the importance of careful work. It was partly for this reason that I suggested so pointedly that you keep a skeptical point of view and look for discrepancies rather than consistency.

Mayall's corroborative evidence, independent in appearance, would weigh more favourably with the jury.

Cross-examination

Another of Hubble's discoveries to face trial was his calculation of great distances to spiral nebulae using the period-magnitude relationship for Cepheid stars. This was a major scientific revolution. Shapley, receiving the news in a letter from Hubble, said "here is the letter that has destroyed my universe"⁴. The focus of research at Harvard and elsewhere abruptly shifted from the local area to distant galaxies. The only evidence against Hubble's great distances was van Maanen's asserted detection of motions in spiral nebulae⁵. "So a consistent picture of the spiral is built on the basis of the Cepheid criterion... save for a disturbing shadow cast by the measures of Mr van Maanen", wrote Hubble in notes for a talk before colleagues at Mount Wilson.

The shadow could be shortened, by stripping away the corroboration claimed by van Maanen⁶. Hubble found only one

[claimed corroborations] as 'similar motions', implying corroborative evidence, indicates an uncritical imagination"⁷.

Additionally, van Maanen claimed corroboration by Seth Nicholson, a colleague at Mount Wilson who had checked some of van Maanen's measurements. Hubble, though, observed that Nicholson's measures gave "results within the uncertainties of the material and hence do not themselves definitely indicate rotation"⁷.

Van Maanen took Hubble's criticism "in a wildly personal manner"⁸. Shapley, a friend of van Maanen, was also disturbed, and felt that Hubble could afford to be less openly critical⁹. At first, Hubble was. Mrs Hubble later remembered that¹⁰:

It was, the contradiction, not very important in the long run. The work had become so obvious and so extensive that van Maanen's measurements could be ignored... Van Maanen had refused to discuss the subject or to re-examine his work. Edwin had decided to ignore the discrepancy for the time and go on with his work. A friend of his told me years afterwards that he [Hubble] said to him 'They asked me to give him [van Maanen] time. Well, I gave him time, I gave him ten years.' He [Hubble] knew the measurements were wrong, his close colleagues knew. To him that would be enough. To publish would be to expose van Maanen to public humiliation.

But the ten years passed and van Maanen's measurements remained. Indeed, they may have begun to draw support once again. As Hubble explained:

When a speaker at the RAS [Royal Astronomical Society] announced, if it were not for van Maanen's measurements, Hubble's results might be accepted, I decided to make the measurements.

Hubble had come to fear loss of support for his work unless van Maanen's contradictory evidence were challenged. Another potential loser was the Mount Wilson Observatory if Hubble and van Maanen were to engage in an unseemly public quarrel. Frederick Seares, editor of Mount Wilson publications, ruled Hubble's prosecution of van Maanen out of order¹¹:

For two men in the same institution there is opportunity for personal contact and for direct examination of each other's results, and hence for the private adjustment of differences in opinion. The institution itself it seems to me, is under obligation to see that all adjustment possible be made in advance of publication. If agreement cannot be attained it may be necessary for the institution to specify how results shall be presented to the public. In that event, however, there must be opportunity for the expression

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Edwin Hubble, photographed by Jay Barrie, from the posthumous *The Nature of Science and Other Lectures* (Huntington Library, San Marino, 1954).

of the corroborations deserving of consideration, and in it the probable error exceeded the mean motion measured. Another study was vitiated by internal inconsistencies and in two others measured motion was in the opposite direction to that reported by van Maanen. Hubble later wrote, in a manuscript never published, that "reference to these

of individual opinion, but any such expression should be concerned only with the scientific aspects of the question at issue.

It seems Hubble was not completely happy with the decision, as Seares added¹¹:

The institution has, I think, the right to enforce this procedure; but in certain cases it may be wiser to waive its technical right and say to a dissatisfied individual, 'Print what you like, but print it elsewhere'.

Hubble needed to rephrase his questions of van Maanen, at least in Seares' courtroom.

Faced with the choice between leaving van Maanen's work unchallenged or challenging it directly in an ungentlemanly manner, Hubble found an ingenious solution. He made new measures and included Nicholson in the effort¹². Thus a former witness for the defence was to give testimony favourable to the prosecution. Nicholson did not explicitly recant his earlier evidence or van Maanen's claimed corroboration, but jurors could easily have formed that impression. Hubble may have hoped they would. And, by using indirect implication rather than direct criticism, Hubble avoided alienating the jury of scientific peers with an explicit attack on a colleague.

In a brief paper immediately following Hubble's, van Maanen in a face-saving gesture presented new measures in partial confirmation of his own while conceding that "it is desirable to view the motions with reserve"¹³. It was more a plea of *nolo contendere* than of innocent or guilty. Obviously the revised pleading had been negotiated with Hubble.

Innocent by implication

In another case Hubble also encouraged jurors to hear more than the witness said, and less. The otherwise puzzling appearance of colleagues ostensibly to verify a calculation scarcely requiring verification becomes understandable as a bit of legalistic legerdemain in aid of the establishment of the velocity-distance relation.

An increase in the red shift or velocity of extragalactic nebulae proportional to their distances from the Solar System had been discussed by the Dutch astronomer Willen de Sitter as early as 1917^{14,15}, and in 1924 Ludvik Silberstein at Rochester had attempted to establish de Sitter's predicted velocity-distance relation using globular clusters^{16,17}. Silberstein, however, had not used all the data available and he had excluded three clusters with velocities not consonant with the theory he was attempting to prove. Strömberg¹⁸ and Lundmark¹⁹ at Mount Wilson found no correlation between the radial velocities and the distances of globular clusters.

They carefully distinguished Silberstein's discredited work from the general hypothesis of a velocity-distance relation. A conclusive test of de Sitter's hypothesis would have to wait a determination of

distances of spiral nebulae because globular clusters might not be distant enough to have a measurable velocity-distance relation. Nevertheless, theoretical preconception such as had influenced Silberstein's analysis was suspect.

Motivated by de Sitter's theory, Hubble had Humason at Mount Wilson systematically observe faint and more distant nebulae to see if they had larger velocities than closer nebulae. Yet in 1929, in the resulting report, Hubble emphasized empirical aspects of his work²⁰. Humason published his measures separately²¹, so Hubble could refer to them as independent evidence. Not until the final paragraph did Hubble mention de Sitter or theory, then noting that the velocity-distance relation might represent the de Sitter effect and might be of interest for cosmological discussion. An understated introduction to the key to the scientific exploration of the universe!

Hubble did emphasize from the beginning paragraph a connection between his velocity-distance relation and earlier determinations of the motion of the Sun with respect to the extragalactic nebulae involving a velocity factor, the *K* term, apparently varying with distance. The *K* term now was largely irrelevant, because Hubble's new data indicated a linear correlation between velocity and distance regardless of whether the velocities were corrected for the solar motion. Intended or not, the *K* term served as a red herring, diverting attention from suspect theoretical considerations. It also provided an opportunity to introduce Strömberg and Lundmark as witnesses.

Hubble used the assumption of a velocity-distance relation to determine the solar motion relative to two groupings of nebulae. The two calculations produced approximately the same result, thus tending to confirm the assumption. Hubble mentioned that Strömberg had checked the general order of the values using different groupings of nebulae. Hubble did not state explicitly that Strömberg accepted the assumption of the velocity-distance relation, though readers could easily have leapt to that implication.

Lundmark could not be brought into court as readily, since he had by now returned to Sweden. But Hubble could cite Lundmark's solution for the solar motion. No matter that it differed from Hubble's solution, or that Lundmark had used a velocity-distance relation different from Hubble's simple linear proportionality. Hubble did not emphasize these differences to the jurors. Nor did Hubble reveal that he generally distrusted Lundmark's work²². What Hubble wanted was Lundmark's appearance on the stand as a friendly witness, not any actual testimony.

Certainly Hubble required neither Strömberg's nor Lundmark's assistance to check the straightforward calculation of

the solar motion. The testimony of Strömberg and Lundmark was not necessary. The presence of Silberstein's critics, though, implied that Hubble's work did not suffer from the same faults.

Hubble's establishment of an empirical velocity-distance relation, although begun in a climate of scepticism and suspicion, proceeded smoothly and quickly. Within two years, Hubble had decisively won over the jury of scientific opinion. The scientific case was outstanding, and it did not hurt to have it argued by a skilled barrister.

Hubble later described modern science in the language of law, writing that science deals "with judgments on which it is possible to obtain universal agreement" and "agreement is secured by observation and experiment—impartial courts of appeal"²³. He made the observations and experiments necessary to win his cases, and skillfully argued the cases.

If Hubble's orchestration of evidence appears unusual, it is because of the conductor's skill and the audience's incomplete view of science. No doubt many scientists orchestrate their evidence; Hubble's orchestration is unusual in the degree to which it has been analysed.

It is not a matter of the fabrication or falsification of data, nor the use of fudge factors. Hubble's data was honest. And if he tilted the playing field upon which his interpretations of data battled others, it often was to restore a balance, to overcome prejudice against his ideas or in favour of others and to attain a fair hearing.

The propriety of any particular orchestration of evidence may be questioned. Perhaps a heroic image of pristine science excludes all orchestration of evidence. But advocacy is an integral part of real science, not to be judged good or bad but to be recognized and studied.

- Hubble, E. to J. Moore, 23 Oct. 1932 and to N. Mayall, 30 Oct. 1933; Hubble Collection, Huntington Library.
- Hubble, E. to N. Mayall, 23 Feb. 1934; quoted in N. Hetherington, *Hist. Stud. Phys. Sci.* 13, 41-67 (1982).
- Hubble, E. to N. Mayall, 23 Feb. 1934.
- Payne-Gaposchkin, C. *An Autobiography and other Recollections* 209 (Cambridge University Press, 1984).
- Hetherington, N. *Nature* 306, 727-730 (1983); *Q. Jl R. astr. Soc.* 13, 25-39 (1972).
- van Maanen, A. *Astrophys. J.* 57, 264-278 (1923).
- Hubble, E. Manuscript, Hubble Collection.
- Hubble, E. to H. Shapley, 11 March 1925; quoted in Smith, R. *The Expanding Universe* (Cambridge University Press, 1982).
- Shapley, H. to F. Seares, 12 May 1925, *ibid.*
- Hubble, Mrs E. to Michael Hoskin, 7 March 1968, Hubble Collection.
- Seares, F. to G. Hale, 24 Jan. 1935; quoted in Smith, R. *The Expanding Universe* 136 (Cambridge University Press, 1982).
- Hubble, E. *Astrophys. J.* 81, 334-335 (1935).
- van Maanen, A. *Astrophys. J.* 81, 336-337 (1935).
- de Sitter, W. *Mon. Not. R. astr. Soc.* 78, 3-28 (1917).
- Hetherington, N. *Isis* 62, 309-313 (1971).
- Silberstein, L. *Phil. Mag.* 47, 907-917; 48, 619-628 (1924), *Mon. Not. R. astr. Soc.* 84, 363-366 (1924).
- Hetherington, N. thesis, Indiana Univ. (1970).
- Strömberg, G. *Astrophys. J.* 61, 353-362 (1925).
- Lundmark, K. *Mon. Not. R. astr. Soc.* 84, 747-769 (1924).
- Hubble, E. *Proc. natn. Acad. Sci. U.S.A.* 15, 168-173 (1929).
- Humason, M. *Proc. natn. Acad. Sci. U.S.A.* 15, 167-168 (1929).
- Smith, R. *The Expanding Universe*, 183 (Cambridge University Press, 1982).
- Hubble, E. *The Nature of Science and Other Lectures* (Huntington Library, San Marino, California, 1954).

Is the galactic centre black hole a dwarf?

David A. Allen* & Robert H. Sanders†

* Anglo-Australian Observatory, PO Box 296, Epping, New South Wales 2121, Australia

† Kapteyn Astronomical Institute, Postbus 800, 9700 AV Groningen, The Netherlands

A 2.2- μm image of IRS 16, the relatively blue infrared source near the galactic centre, is presented at the highest spatial resolution yet reported. It comprises several discrete, extended sources. By intercomparison of radio and infrared images of the region, the non-thermal radio source Sgr A can be located relative to IRS 16 without the uncertainty of intermediate astrometric frames. No component of IRS 16 is coincident with Sgr A*. These data preclude the presence of a massive ($>10^6 M_\odot$) black hole at the galactic centre. Any black hole present has a mass nearer to $100 M_\odot$.*

EVIDENCE for a black hole at the galactic centre has been discussed by Brown and Liszt¹ and is summarized here. The primary evidence is the presence of a compact radio source, Sgr A*, first described by Balick and Brown². Its linear size is <20 AU (ref. 3) whilst its radio luminosity, $2 \times 10^{34} \text{ erg s}^{-1}$, is many orders of magnitude greater than that of any pulsar. Within the Galaxy, Sgr A* is unique. Probably coincident with it is a soft X-ray source similar to those in active galactic nuclei but emitting only $10^{35} \text{ erg s}^{-1}$. Moreover, within 4° of the centre is an extraordinary, variable source emitting the positronium annihilation line at 0.5 MeV (ref. 4).

Around Sgr A* is a striking radio continuum structure—the ‘mini-spiral’ of Ekers *et al.*⁵—which suggests special kinematic activity, either mass ejection¹ or accretion of small clouds onto a compact mass⁶. The inner region is largely free from dust, but far-infrared maps⁷ indicate a ring of dust of inner radius about 1 pc which may coincide with portions of the mini-spiral. This dust is heated by a central source of radiation with a total optical/ultraviolet luminosity $\sim 10^{40} \text{ erg s}^{-1}$ and a temperature $<35,000 \text{ K}$ (ref. 8).

At spatial resolutions of 1–2 arc s, attention is focused on the near-infrared source IRS 16 (ref. 9), which lies very close to Sgr A*. Near IRS 16, emission lines of H and He I include a broad component^{10,11} of velocity width $1,500 \text{ km s}^{-1}$. Assuming virial

equilibrium, the dynamical mass near IRS 16 is thus $10^7 M_\odot$. IRS 16 has been resolved¹² into three merged components aligned roughly along the galactic equator.

It has been recognized for many years¹³ that the near-infrared emission from the central region of the Galaxy traces the stellar density distribution. The implied distribution approximates $r^{-1.8}$ over the range $0.1 < r < 40 \text{ pc}$ (refs 14, 15). Because the exponent is tantalizingly close to that predicted for the cusp around a black hole, it is tempting to identify IRS 16 with the bright central region of the cusp. If a massive black hole is present, we therefore expect a spatial coincidence between Sgr A* and one component of IRS 16.

As a test of the coincidence, astrometry of IRS 16 has been performed relative to a system of optical standard stars. Sgr A* is also located relative to the same reference frame of standard stars through a number of compact radio sources with optical counterparts. The accumulated uncertainty in the infrared–radio astrometry is estimated to be 0.5 arc s. The two determinations^{12,16} suggest that IRS 16 centre coincides with Sgr A*, although the radio source appears displaced westwards according to the data of ref. 12.

A second, potentially more accurate, technique is that of directly comparing the radio and infrared maps and positioning them by common features. The first such comparison was made

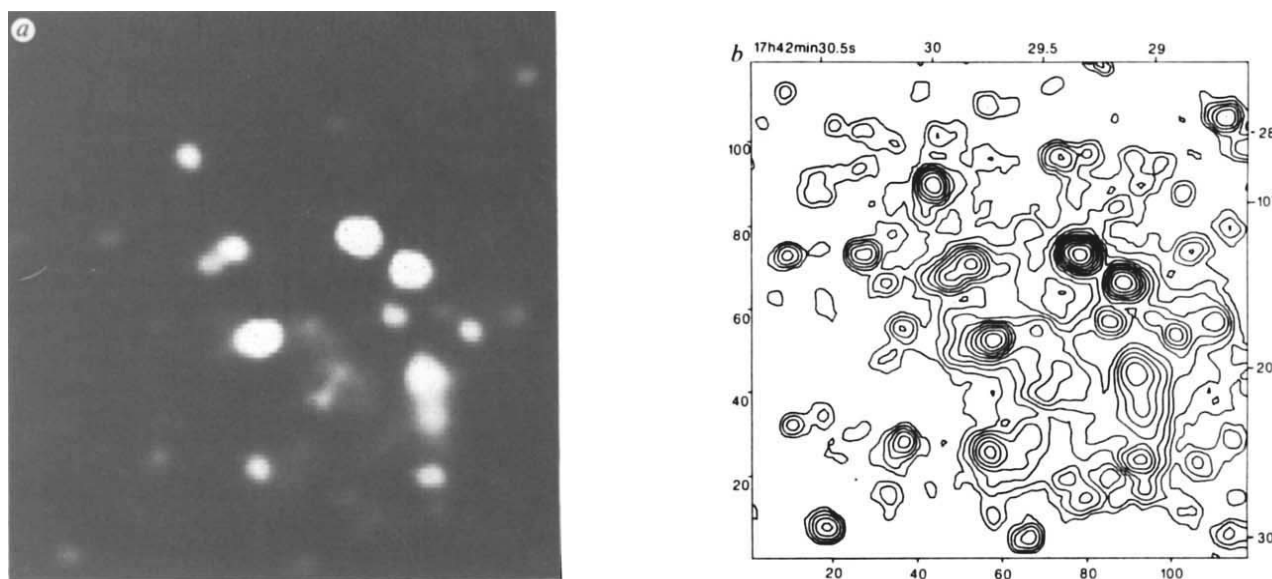


Fig. 1 *a*, The galactic centre at $3.8 \mu\text{m}$; image made on 11 July 1984. The image measures 29.5 arc s on a side. IRS numbers are given for some sources. North is up and east to the left on this and all subsequent images. IRS 16 is the complex of sources to the west of IRS 1W. *b*, The same image in contour form. Contours are separated by 0.5 stellar magnitudes. A point source of $3.8 \mu\text{m}$ (L') magnitude 4.5 will just reach the highest contour; one of magnitude 9.5 the lowest plotted.



Fig. 2 A reproduction of the 6-cm data of Lo and Claussen⁶. Crosses mark the centres of the compact infrared sources as found by the best fit of the infrared data.

by Ekers¹⁷, who used a published map of Br γ emission¹² to conclude that Sgr A* lies about 1 arc s west of every component of IRS 16.

We present an improved comparison of the radio and infrared maps of the galactic centre using new data at 3.8 and 4.8 μm . In agreement with Ekers, we find that Sgr A* does not correspond to the bright components of IRS 16.

Infrared data and radio structure

At short wavelengths, the galactic centre region is dominated by highly reddened stars and IRS 16 is a prominent object^{9,18}. At longer wavelengths the pattern of radiation approaches that of the thermal radio, but IRS 16 cannot be detected. To make a direct comparison with the radio, we made a deep image of the region at 3.8 μm . Data described by Storey and Allen¹² have demonstrated that IRS 16 has identical structure at 1.6, 2.2 and 3.8 μm (a colour composite of the three was constructed and reproduced in ref. 19). The new image is reproduced in Fig. 1. We used the 3.9-m Anglo-Australian Telescope, with the infrared photometer-spectrometer operated in its imaging mode. At the focal plane a 1-arc s circular aperture was used; the seeing did not appreciably degrade the resolution. The image was sampled every 0.5 arc s, and in constructing Fig. 1 we have expanded the data to 0.25 arc s pixels by linear interpolation. Four separate images were combined into Fig. 1, allowing us to assess the internal positional accuracy as better than 0.15 arc s.

For the comparison with the radio data, we used the image at 6 cm made by Lo and Claussen⁶. This image (Fig. 2) also has a spatial resolution of about 1 arc s. It was sent in digital form to permit manipulations of the two images using Starlink software on the AAO VAX 11/780.

On the 3.8- μm image, the density of point sources is high. Some of them have not been recorded at other wavelengths. Most lack radio counterparts, but several diffuse patches of emission do correspond well with the radio map, and there is undoubtedly a strong correlation in the vicinity of IRS 2/13, on which we place much reliance. Nonetheless, before a secure fit of the two images could be made, a map at a longer wavelength was necessary to locate the compact sources IRS 1W, 5 and 10 relative to the northeasterly arm of radio emission.

An image at 4.8 μm was secured at a later date. The observational configuration duplicated that of the 3.8- μm map, but poorer seeing degraded the spatial resolution. Figure 3 repro-

duces this image. The details resemble more closely those at 6 cm, but with local enhancements of intensity. IRS 2/13 remains prominent. IRS 1W, 5 and 10 are now seen to lie east of the radio structure. IRS 1W and 10 have weak counterparts in the radio, whilst IRS 5 barely contacts the radio contours. Their locations strongly suggest a relationship with the mini-spiral. Presumably, they are single or multiple stars close enough to the gas clouds to cause local heating of the dust. Their displacement from the ionized gas is itself an interesting matter not addressed here.

With the help of the 4.8- μm image we located the 3.8- μm and 6-cm images with a relative uncertainty little more than 1 pixel or 0.25 arc s. Sgr A* was then found to lie significantly to the west of all components of IRS 16. Attempts to overlay the images with Sgr A* superimposed on any component of IRS 16 do not permit satisfactory registration of the infrared and radio sources.

IRS 16 in detail

Our attention turned to the region west of IRS 16. A small strip image was made at 2.2 μm (K) using a focal-plane diaphragm of 0.7 arc s diameter. Seeing at that wavelength was mostly below 0.5 arc s at the time. Several images were made, but only those with consistently good seeing were combined to create Fig. 4, which has almost double the spatial resolution of previous data¹². The strip map includes IRS 7, and some faint point sources to the south. IRS 16 NE is bisected at the left edge of the image.

IRS 16 SW and centre are now seen to be discrete sources extended on the scale of about 1 arc s. A previously ignored component, IRS 16 NW, is unresolved, and for this reason appears as intense as the peak of SW and centre. Faint emission is seen west of IRS 16 centre, at the location of Sgr A*. Adopting $K = 6.7$ for IRS 7, the various components of IRS 16 peak at about 10.0 mag in the 0.7-arc s beam. The sources at the location of Sgr A* has $K = 11.4$. Both of these numbers have been corrected from the raw data for the fact that the darkest part of Fig. 4 (relative to which the photometry is performed) lie above the dark sky level on the data of ref. 12.

We now examine the impact of these observations on models invoking a massive black hole at the site of Sgr A*.



Fig. 3 4.8- μm image (8 November 1984). Although inferior to Figs 1 and 2, features of both these images can be seen.

IRS 16 as the cusp core

We adopt the popular assumption that Sgr A* is a 10^6 – $10^7 M_\odot$ black hole. We first ask whether the observed structure of IRS 16 is consistent with this complex source being the core of a cusp around the black hole.

A point mass in an extended distribution of stars dominates the gravitational potential within the classical accretion radius (r_a):

$$r_a = \frac{2GM_h}{v_s^2} = \frac{M_h}{5 \times 10^6 M_\odot} \left[\frac{225 \text{ km s}^{-1}}{v_s} \right]^2 \text{ pc} \sim 20 \text{ arc s} \quad (1)$$

where M_h is the mass of the object, and v_s is the stellar velocity dispersion. Due to energy diffusion, a cusp in the stellar density distribution of the form $\rho \propto r^{-1.75}$ will develop^{20,21} within r_a on a timescale longer than the initial two-body relaxation time²²:

$$t_R = \frac{\sqrt{2} v_s^3}{\pi G^2 M_*^2 n \ln(0.5N)} \quad (2)$$

in a stellar system with a mean stellar mass M_* , a stellar density of n , and a total number of stars N . The inferred properties of the inner 1 pc of the galactic nucleus ($n \sim 4 \times 10^5 \text{ pc}^{-3}$, $v_s \sim 200$ – 300 km s^{-1})¹⁴ imply that $t_R \sim 10^{11} \text{ yr}$ at the present epoch. However, even without energy diffusion, the adiabatic growth of a black hole produces a cusp in the stellar density²³ of the form $r^{-1.5}$. Therefore, the formation of a cusp around such a hypothetical black hole in the galactic centre would seem inevitable.

The observed density index in the central parsec of the Galaxy is -1.8 , consistent with that of an energy diffusion cusp. Using existing data^{14,15}, the integrated magnitude at K within a circular aperture of radius r is $m(r) = 8.5 - 3 \log(r)$. Within the 0.7-arc s aperture of the present observation, we should therefore expect $K = 9.9$, consistent with any of the IRS 16 components, but 1.5 mag brighter than found at the location of Sgr A*.

Can the hole migrate away from the centre of its cusp due to interactions with cusp stars²⁴? Although the hole may deviate from the centre-of-mass of the hole-cusp system, it cannot deviate significantly from the position of maximum brightness of the cusp. The innermost, tightly bound stars will follow the hole on its brownian motion resulting from interactions with more distant cusp stars. The only possible deviation of the hole from the point of maximum brightness can be due to statistical fluctuations in the space density of cusp stars, and this does not become significant except at radii within which there are only about 10 cusp stars, in this case 10^{-5} pc . We find a displacement of Sgr A* from the IRS 16 components of $>1,000$ times this limit, implying that none of the bright IRS 16 components can be the core of a hypothetical cusp.

Association of young stars

The fact that Sgr A* does not coincide with any of the bright components of IRS 16 implies that a possible black hole associated with this object does not emit significantly at $2.2 \mu\text{m}$. If we characterize the radiation from the hole's accretion disk by a black body, we have an upper limit to its temperature of $35,000 \text{ K}$ from the presence of Ne II (ref. 8). To provide the optical and ultraviolet energy seen to be heating the dust requires a K magnitude of about 8, which is clearly precluded. This is a lower bound to K for any reasonable energy distribution other than line emission. Thus no object associated with Sgr A* provides the bulk of the luminosity of the galactic centre. The observed K magnitude of 11.4 implies an upper limit to the luminosity of about $10^{39} \text{ erg s}^{-1}$, which is compatible with the low level of radio and X-ray activity, and with the presence of an inactive, massive black hole there. But then, what is the source of ionizing radiation at the galactic centre?

There are many discrete sources in the galactic centre region which must contain stars of significant luminosity. The differences between the infrared and radio continua alone indicate

this fact. The discrete infrared sources can be interpreted either as local centres of heating²⁵ or as external heating of local density enhancements²⁶. Most of these sources appear to suffer local extinction from the dust that re-radiates at long wavelength; they cannot illuminate the dust ring 2 pc from the centre because it is transparent to their radiation. We should seek the required luminosity in the bluer infrared sources. IRS 16 is the brightest of these.

Geballe *et al.*¹¹ demonstrated that IRS 16 (loosely defined within a 4-arc s beam) is responsible for most of the ionization within the central parsec. Its integrated K magnitude is about 8, so if the constituent stars approach $35,000 \text{ K}$ in temperature, they alone can account for the required luminosity. The concentration of M supergiants near the galactic centre may be explained by a burst of star formation which occurred in the nuclear region $\sim 10^7 \text{ yr}$ ago^{27,28}. Such a burst would leave at least one concentration of B stars now. We therefore regard all the components of IRS 16 as clusters of hot stars, remnants of the last starburst. If their mean spectral type is B0, about 100 are required, or roughly 25 in each of the four major components. The clumpy, irregular appearance of these components is better explained in this way than as the more symmetrical cusp or compact accretion disk associated with an active black hole.

However, the proximity of these objects to the point radio source severely constrains the mass of the black hole. If, as we suggest, the sources are unbound clumps of B stars, then they will disperse on a timescale roughly comparable to the orbit time about the black hole, or

$$t_D \sim 2 \left(\frac{GM_h}{r_0^3} \right)^{-1/2} \sim 500 \left(\frac{M_h}{5 \times 10^6 M_\odot} \right)^{-1/2} \text{ yr} \quad (3)$$

where $r_0 \sim 0.1 \text{ pc}$ is the typical projected distance of the clumps from the hole. The probability of observing two or three such

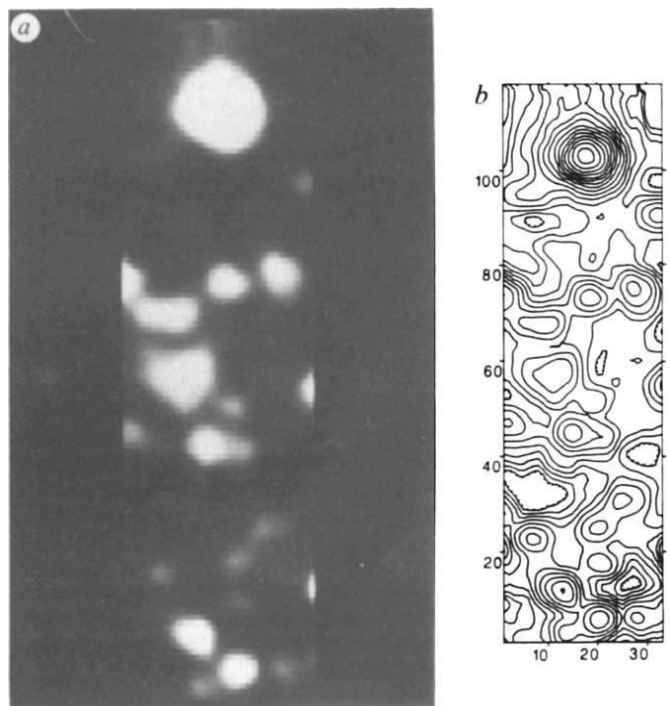


Fig. 4 *a*, Strip image at $2.2 \mu\text{m}$ through IRS 7 and 16. The image measures $5.0 \times 17.7 \text{ arc s}$ and the pixels are 0.15 arc s square. The spatial resolution, about 0.8 arc s , is a factor of two higher than any previous data. The IRS sources are identified and Sgr A* is marked by a cross of dimension 0.5 arc s^{-1} , the 2σ uncertainty in its position. *b*, The same image in contour form. The contours are separated by 0.5 stellar magnitude and range between 6.5 and 14.0 in K magnitude.

clumps in the central 0.1 pc is extremely low unless star formation is occurring continuously. For any likely initial mass function, this is precluded by the cool radiation field of the region. Moreover, the tidal field of a massive black hole would require implausibly dense protostellar clouds:

$$\rho > \frac{6M_h}{r_0^3} \sim 7 \times 10^{-13} \frac{M_h}{5 \times 10^6 M_\odot} \text{ g cm}^{-3} \quad (4)$$

What alternative interpretations of the IRS 16 sources are possible?

(1) They are background or foreground sources seen in projection within 2 arc s of the black hole. Not only would it be highly fortuitous that the only blue objects seen in the inner 2 pc apparently lie so close to the centre, but we should have to find another source of luminosity to power the far-infrared emission.

(2) The clumps are dense gravitationally bound subclusters. It has been suggested²⁹ that globular clusters will be dragged into the centre of a galaxy by dynamical friction. Again, to prevent tidal disruption, their mean density must be given by equation (4). This corresponds to a stellar density in the subcluster of almost $10^{10} M_\odot \text{ pc}^{-3}$ within 0.02 pc. Certainly, no presently observed globular cluster cores have these properties; in the most dense globulars $n \sim 10^5 M_\odot \text{ pc}^{-3}$ (ref. 30).

The clumps could be the highly evolved cores of old, massive globular clusters. But the present stellar relaxation time in such systems is only $\sim 10^7$ yr. Because the lifetime of such a cluster is no more than $30 t_R$ (ref. 31), to see several of them they would have to be replaced or reformed every 10^8 yr. External replacement over 10^{10} yr would have produced an unacceptably large mass in the central 0.1 pc of the Galaxy, exceeding $3 \times 10^7 M_\odot$.

(3) Several massive black holes occupy the central regions, each with its own cusp of bound stars. Such a configuration is dynamically unstable³², leading to the formation of at most a binary system. Nor does it explain the uniqueness of Sgr A*.

(4) The bright components are one or more background stars in the Galaxy which have been gravitationally imaged by the massive black hole. The required amplification factor, 10^5 – 10^6 , implies that the star lies within 4×10^{-5} arc s of the direction of the black hole. Given the mean density of stars along a line-of-sight through the Galactic plane, the probability of this is $\sim 10^{-7}$.

A low-mass black hole

We have attempted to reconcile the observations with the existence of a black hole coincident with Sgr A* and of mass exceeding $10^6 M_\odot$ and have failed to do so. Our conclusion, which we now justify, is that the black hole within Sgr A* is no more massive than about $100 M_\odot$.

This alternative does not contradict the great body of observational material. None of the phenomena noted above requires a black hole more massive than $100 M_\odot$. Suggestions³³ that the

kinematic evidence requires a point mass $> 10^6 M_\odot$ assume that the gas motions within 0.5 pc of the centre are virialized and thus trace the gravitational potential. This need not be the case; indeed, the broad emission lines seen in the direction of IRS 16 may be high-velocity outflow from the region of a low-mass black hole (as suggested in ref. 11). The fact that the central ionization source can be attributed to the clumps of B0 stars identified with IRS 16 removes the justification for a more bizarre source. Moreover, the most straightforward interpretation³⁴ of the positronium annihilation line implies a black hole of mass no more than a few hundred $M_\odot$.

The similarity of the stellar density distribution in the inner parsec to that in a cusp around a massive black hole is not compelling: the density law also resembles an isothermal sphere (r^{-2}) with a small (0.1 pc) core and a central density $> 10^7 M_\odot \text{ pc}^{-3}$. The two-body relaxation time in that core would be 2×10^9 yr and the stellar collision time would be 9×10^9 yr. As the timescale for dynamical evolution is the smaller of t_{coll} and 20 – $30 t_R$, the structure need change very little over the age of the Universe.

The dispersal of unbound groups of young stars would be considerably less rapid in this case. In a constant-density or 'harmonic' case, where all orbit periods are equal, structure would never disappear. It is unrealistic to assume that the IRS 16 associations lie in a pure harmonic oscillator potential; a 5–10% variation in the density of the underlying isothermal core would cause dispersal by phase mixing in 10–20 orbit times, or about 10^4 yr. Moreover, if $M_h < 5 \times 10^4 M_\odot$, the Roche limit constraint (equation (4)) is relaxed and the associations could be gravitationally bound subclusters of no more than $1,000 M_\odot$. Thus the existence of associations or clusters of young stars, as evidenced by the structure in IRS 16, is only plausible if the mass of the hole is considerably less than the mass of the isothermal core, that is, $M_h < 10^5 M_\odot$.

Determination of the velocities of the IRS 16 components would verify (or falsify) our claim that the black hole is much less massive than $10^6 M_\odot$. For a massive black hole, the IRS 16 components should have a velocity $> 450 \text{ km s}^{-1}$. This could be detected in the radial velocity of the associated Br γ line.

The arguments we have given do not tightly constrain the mass of the black hole. However, consideration of the growth rate of a black hole in the stellar density of the galactic centre make it unlikely that the mass is as high as $1,000 M_\odot$; such a black hole would consume the galactic core, thus growing to a mass of 10^6 – $10^7 M_\odot$ in 10^9 yr (refs 35, 36). The upper bound to the luminosity of Sgr A* is about that of a $100 M_\odot$ black hole accreting at the Eddington luminosity. Therefore, we conclude that the central parsec of the Galaxy comprises an isothermal sphere of stars having a core radius ~ 0.1 pc and central density $> 10^7 M_\odot \text{ pc}^{-3}$, a few small associations or clusters of newly formed B stars within that core, and an accreting black hole of mass $\leq 100 M_\odot$.

Received 29 July; accepted 21 October 1985.

1. Brown, R. L. & Liszt, H. L. *A. Rev. Astr. Astrophys.* **22**, 223–265 (1984).
2. Balick, B. & Brown, R. L. *Astrophys. J.* **194**, 265–270 (1974).
3. Lo, K. Y. *et al. Nature* **315**, 124–126 (1985).
4. Riegler, G. R. *et al. Astrophys. J. Lett.* **248**, L13–L16 (1981).
5. Ekers, R. D., van Gorkom, J. H., Schwartz, U. J. & Goss, W. M. *Astr. Astrophys.* **122**, 143–150 (1983).
6. Lo, K. Y. & Claussen, M. J. *Nature* **306**, 647–651 (1983).
7. Becklin, E. E., Gatley, I. & Werner, M. W. *Astrophys. J.* **258**, 135–142 (1982).
8. Lacy, J. H., Townes, C. H., Geballe, T. R. & Hollenbach, D. J. *Astrophys. J.* **241**, 132–146 (1982).
9. Becklin, E. E. & Neugebauer, G. *Astrophys. J. Lett.* **200**, L71–L74 (1975).
10. Hall, D. N. B., Kleinmann, S. G. & Scoville, N. Z. *Astrophys. J. Lett.* **260**, L53–L57 (1982).
11. Geballe, T. R. *et al. Astrophys. J.* **284**, 118–125 (1984).
12. Storey, J. M. V. & Allen, D. A. *Mon. Not. R. astr. Soc.* **204**, 1153–1161 (1983).
13. Becklin, E. E. & Neugebauer, G. *Astrophys. J.* **151**, 145–161 (1968).
14. Bailey, M. E. *Mon. Not. R. astr. Soc.* **190**, 217–225 (1980).
15. Allen, D. A., Hyland, A. R. & Jones, T. J. *Mon. Not. R. astr. Soc.* **204**, 1145–1152 (1983).
16. Henry, J. P., De Poy, D. L. & Becklin, E. E. *Astrophys. J. Lett.* **285**, L27–L30 (1984).
17. Ekers, R. D. *Pap. at Workshop Galactic Centre* (California Institute of Technology, city? 1984).

18. Bailey, J., Hough, J. H. & Axon, D. J. *Mon. Not. R. astr. Soc.* **208**, 661–671 (1984).
19. Henbest, N. *New Scientist* **97**, 460–461 (1983).
20. Bahcall, J. N. & Wolf, R. A. *Astrophys. J.* **209**, 214–232 (1976).
21. Duncan, M. J. & Shapiro, S. L. *Astrophys. J.* **268**, 565–581 (1983).
22. Spitzer, L. & Harm, R. *Astrophys. J.* **127**, 544–550 (1958).
23. Young, P. J. *Astrophys. J.* **217**, 287–295 (1977).
24. Lin, D. N. C. & Tremaine, S. *Astrophys. J.* **242**, 789–798 (1980).
25. Lebofsky, M. J., Rieke, G. R., Deshpande, M. R. & Kemp, J. C. *Astrophys. J.* **163**, 672–676 (1982).
26. Gezari, D. Y. *et al. NASA techn. Mem.* 86225 (1985); *Astrophys. J.* (in the press).
27. Lebofsky, M. J., Rieke, G. R. & Tokunaga, A. T. *Astrophys. J.* **263**, 736–740 (1982).
28. Lacy, J. H., Townes, C. H. & Hollenbach, D. J. *Astrophys. J.* **262**, 120–134 (1982).
29. Tremaine, S. D., Ostriker, J. P. & Spitzer, L. *Astrophys. J.* **196**, 407–411 (1975).
30. Peterson, C. J. & King, I. R. *Astr. J.* **80**, 427–436 (1975).
31. Lightman, A. P. & Shapiro, S. L. *Rev. mod. Phys.* **50**, 437–481 (1978).
32. Saslaw, W. C., Valtonen, M. J. & Aarseth, S. J. *Astrophys. J.* **190**, 253–270 (1974).
33. Crawford, M. K. *et al. Nature* **315**, 467–470 (1985).
34. Lingenfelter, R. E. & Ramaty, R. *The Galactic Center* 148–159 (American Institute of Physics, New York, 1982).
35. Hills, J. G. *Nature* **254**, 295–298 (1975).
36. Ozernoy, L. M. *IAU Colloq.* **45**, 121–124 (1977).

Relationship between atmospheric CO₂ variations and a satellite-derived vegetation index

C. J. Tucker*, I. Y. Fung†, C. D. Keeling‡ & R. H. Gammon§

* NASA/Goddard Space Flight Center, Code 623, Greenbelt, Maryland 20771, USA

† NASA/Goddard Institute for Space Studies, New York, New York 10025, USA and Lamont-Doherty Geological Observatory of Columbia University, Palisades, New York, New York 10964, USA

‡ Scripps Institution of Oceanography, La Jolla, California 92093, USA

§ NOAA/GMCC, Boulder, Colorado 80302, USA

Estimates derived from satellite data of photosynthetically active radiation absorbed by terrestrial vegetation correlate with atmospheric carbon dioxide concentrations measured at surface recording stations, suggesting that satellite data can be used to estimate terrestrial photosynthesis.

TERRESTRIAL photosynthesis is inversely related to atmospheric carbon dioxide concentrations and is an important component of the global carbon cycle. The lack of temporal resolution and global data has prevented comparisons between estimates of terrestrial photosynthesis and measurements of atmospheric CO₂ concentrations from remote maritime monitoring stations. This is now possible using data derived from meteorological satellites.

Primary production

The biosphere absorbs CO₂ from the atmosphere through photosynthesis which incorporates absorbed carbon into carbohydrates, giving off O₂. In the terrestrial biosphere, CO₂ absorption occurs in the green leaves of plants, where intercepted photosynthetically active radiation (IPAR) (0.4–0.7 µm) is absorbed by the photosynthetic pigments. The energy of the absorbed photons is used to drive the reaction of photosynthesis, yielding carbohydrates through primary production. Incident radiation in the 0.4–0.7-µm region is strongly absorbed by green leaves (by scattering and absorption by photosynthetic pigments), and radiation in the 0.7–1.1-µm region is strongly backscattered or reflected (by scattering in the absence of absorption)^{1–5}. Thus, reflected radiance from the red and near-infrared spectral regions can be used to estimate the green-leaf biomass of plant canopies^{6,7}, are directly related to the IPAR^{8–13}, the photosynthetic capacity and the resistance to water-vapour transfer of plant canopies¹³.

Atmospheric CO₂

Measurements of atmospheric CO₂ concentrations at monitoring stations have a seasonal oscillation superimposed on an annually increasing trend. The annual amplitude of this oscillation is ~15 parts per 10⁶ (p.p.m.) at Point Barrow, Alaska, and decreases southward to ~6 p.p.m. at Mauna Loa, Hawaii, and to ~1–2 p.p.m. in the Southern Hemisphere. The annual maximum atmospheric CO₂ concentration in the Northern Hemisphere occurs in the late spring or early summer and the annual minimum concentration occurs in the autumn^{14,15}. The annual oscillation of atmospheric CO₂ is due primarily to seasonal exchange of CO₂ with the terrestrial biosphere^{14,16,17}; in contrast, there is little seasonal exchange of CO₂ with the upper ocean^{18,19}. The CO₂ exchange with the biosphere has been simulated using diffusion, advection-diffusion, three-dimensional tracer models^{20–22} and biospheric-exchange functions. Bolin and Keeling²³ and Pearman and Hyson²⁴ derived CO₂ sources and sinks

consistent with atmospheric CO₂ observations, while Azevedo²⁵ constructed simple biospheric-exchange functions based on assumed vegetation seasonality and geographical location. However, the biospheric CO₂ drawdown and release functions that correlated best with measured CO₂ concentrations showed differences by at least a factor of 2 (refs 20–25). The release of CO₂ from terrestrial vegetation can be approximated if primary production and the temperature regime are known; the drawdown of CO₂ by terrestrial vegetation can be determined if primary production is known. Thus, elucidation of the role of the terrestrial biota in the global carbon cycle requires an estimate of the terrestrial primary production; this can be achieved through remote sensing by satellite.

Remote sensing

By monitoring global changes in the IPAR of the terrestrial biota, an estimate of CO₂ drawdown and primary production can be obtained through the relationships between leaf structure, leaf density and light in plant canopies, and that between IPAR and photosynthetic capacity¹³.

The CO₂ cycle involves not only photosynthesis and respiration, but also decomposition of carbon compounds, uptake and release by oceans, combustion of fossil fuels, and fires involving vegetation. Deforestation contributes towards atmospheric CO₂ through burning and/or subsequent decomposition of phytomass, litter and associated carbon concentrations in the soil, although the magnitude of this contribution is uncertain^{26,27}. Remote sensing by satellite enables the temporal and spatial variations in the IPAR of the terrestrial biota to be quantified.

Remote-sensing studies of the biota have used Landsat and meteorological satellites data. Landsat multispectral scanner (MSS) data, with a spatial resolution of 80 m, are not appropriate for the study of seasonal changes in the terrestrial biota, because ~5,000 Landsat MSS scenes would be required to cover the world's surface. The polar-orbiting meteorological satellites of the National Oceanic and Atmospheric Administration (NOAA) are suitable for observing terrestrial IPAR dynamics. The advanced very-high-resolution radiometer (AVHRR) sensor on board these polar-orbiting satellites daily collects global radiance data at a 4-km spatial resolution. Combinations of channel 1 (0.55–0.68 µm) and channel 2 (0.73–1.1 µm) data from the NOAA satellites' AVHRR have been used to study continental land cover, global vegetation phenology and net primary production by grasslands using data from 1 to 4-yr periods^{28–31}.

Data methods and use

Data on monthly atmospheric CO₂ concentrations for 1982–84 were obtained for Point Barrow, Alaska (147° W, 71° N), and

§ Present address: NOAA/PMEL, Seattle, Washington 98115, USA.

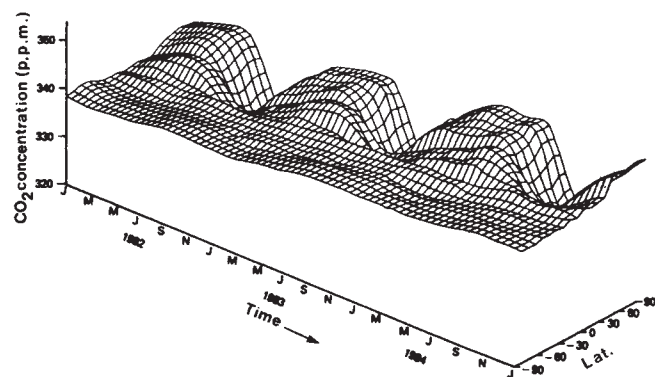


Fig. 1 Variation of global atmospheric CO₂ concentrations with latitude and time based on the NOAA/GMCC flask measurements for 1982–84.

Mauna Loa, Hawaii (156° W, 20° N); and from January 1982 to September 1984 for the South Pole (90° S). Stations there are operated by the NOAA Geophysical Monitoring for Climatic Change (GMCC) programme, and the CO₂ flask samples are analysed at the Scripps Institute of Oceanography. CO₂ concentrations for 20 globally dispersed monitoring sites operated and analysed by NOAA/GMCC for 1982–84 were also obtained.

The atmospheric CO₂ concentration is the combined response to terrestrial biospheric, oceanic and anthropogenic sources and sinks, convolved with long-range transport. Figure 1 shows the total CO₂ signal from January 1982 to December 1984 measured at the 20 global monitoring stations. An annually increasing trend of ~1.5 p.p.m. yr⁻¹ from anthropogenic sources, and an

annual cycle, with an amplitude greatest around 50°–70° N, is evident. To extract the seasonal variation from the CO₂ records, we computed the 12-month running means for each station and subtracted these from the CO₂ record (see ref. 15).

Global weekly data from NOAA-7's AVHRR were obtained for the period from 12 April 1982 to 28 October 1984 (ref. 32). These data were re-mapped from a polar stereographic projection to a Mercator projection so that the world could be represented in a single image. The normalized difference vegetation index (NDVI) $[(C_2 - C_1)/(C_2 + C_1)]$ was calculated^{6,7}, and the highest monthly NDVI value was selected from the weekly data for each month between April 1982 and October 1984 (Table 1). This procedure minimizes the effects of Sun angle, off-nadir viewing, atmospheric pathlength and aerosols, and clouds, all of which decrease the NDVI³³. In winter conditions, spurious NDVI values result from a combination of low solar illumination levels, differing dark-current levels in channels 1 and 2, and pathlength-associated variation in atmospheric effects. These effects are apparent not only during the respective hemisphere winters, but also if the NDVI is calculated from AVHRR data acquired during the night. To compensate for these spurious values, mean monthly surface temperature data were acquired³⁴, and areas within the 0 °C or colder regions were assigned NDVI values of 0.00 because physiological evidence suggests that the rate of photosynthesis at 0 °C is close to zero³⁵.

The monthly NDVI data, which have an average grid-cell size of ~40 km, were used to compute average values for global latitude zones. The zonal averages were multiplied by the land area within each zone (to weight the averages by the land area involved) and summed for various latitude zones; the sum for each larger zone was then divided by the total land area in that zone. The NDVI for a latitude zone represents the potential IPAR; hence, we have assumed that the NDVI is directly related

Table 1 Weekly data used for monthly NDVI composite images

Month	Year	Weeks used	Comments
April	1982	12–18, 19–25, 26–30 April; 1–2 May	Includes 1–2 May
May	1982	3–9, 10–16, 17–23, 24–30 May	
June	1982	31 May; 1–6, 7–13, 14–20, 21–27 June	Includes 31 May
July	1982	5–11, 12–18, 19–25, 26–31 July; 1 August	Includes 1 August
August	1982	2–8, 9–15, 16–22, 23–29 August	
September	1982	30–31 August; 1–5, 6–12, 13–19, 20–26 September	Includes 30 and 31 August
October	1982	4–10, 11–17, 18–24, 25–31 October	
November	1982	1–7, 8–14, 15–21, 22–28 November	
December	1982	29–30 November; 1–5, 6–12, 13–19, 20–26 December	Includes 29 and 30 November
January	1983	3–9, 10–16, 17–23, 24–30 January	
February	1983	31 January; 1–6, 7–13, 14–20, 21–27 February	Includes 31 January
March	1983	28 February; 1–6, 7–13, 14–20, 21–27 March	Includes 28 February
April	1983	4–10, 11–17, 18–24, 25–30 April; 1 May	Includes 1 May
May	1983	2–8, 9–15, 16–22, 23–29 May	
June	1983	30–31 May; 1–5, 6–12, 13–19, 20–26 June	Includes 30 and 31 May
July	1983	4–10, 11–17, 18–24, 25–31 July	
August	1983	8–14, 15–21, 22–28 August	No data 1–7 August
September	1983	5–11, 12–18, 19–25, 26–30 September; 1–2 October	Includes 1 and 2 October
October	1983	3–9, 10–16, 17–23, 24–30 October	
November	1983	31 October; 1–6, 7–13, 14–20, 21–27 November	Includes 31 October
December	1983	5–11, 12–18, 19–25, 26–31 December; 1 January	Includes 1 January
January	1984	2–8, 9–12, 16–22, 23–29 January	No data 13–15 January
February	1984	30–31 January; 1–5, 6–12, 27–28 February; 1–4 March	Includes 30–31 January and 1–4 March; no data 13–26 February
March	1984	5–11, 12–14, 19–25 March	No data 14–18, 26–31 March
April	1984	2–8, 9–14, 16–21, 23–29 April	No data 15 and 22 April
May	1984	30 April; 1–6, 7–13, 14–20, 21–27 May	Includes 30 April
June	1984	4–10, 11–17, 18–24, 25–30 June; 1 July	Includes 1 July
July	1984	2–8, 9–15, 16–22, 23–29 July	
August	1984	30–31 July; 1–5, 6–12, 13–19, 20–26 August	Includes 30 and 31 July
September	1984	3–9, 10–16, 17–23, 24–30 September	
October	1984	1–7, 8–14, 15–21, 22–28 October	

The maximum monthly value was selected from the weekly data.

Fig. 2 Weighted NDVI data plotted against time and latitude zone. Note the highly seasonal effects in the northern latitudes, the influence of deserts in the 20°–30° N latitude zone, the generally constant response in equatorial areas, and the influence of the low proportion of land area south of 30° S. See also Fig. 1. The average NDVI value for each 5° latitude zone has been multiplied by the land area in the zone to compensate for global variations in land area.

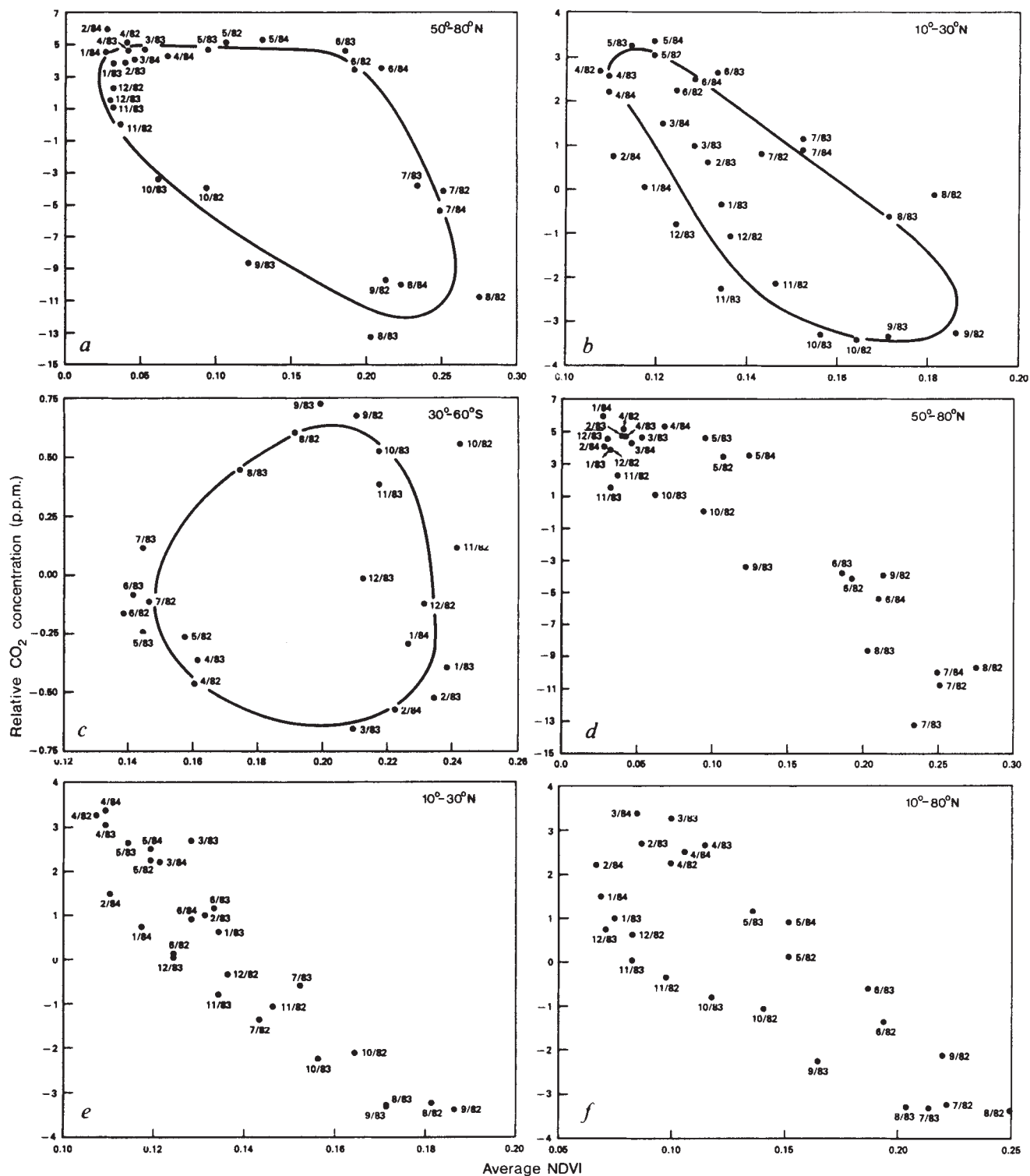
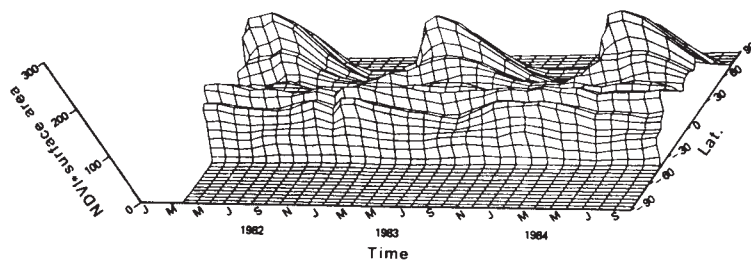


Fig. 3 Relationships between atmospheric CO_2 concentrations and NDVI averages with and without time lags between NDVI measurements and CO_2 measurements. The CO_2 data are from stations at Point Barrow, Alaska: a, no time lag; d, 1-month lag; Mauna Loa, Hawaii: b, no time lag; e, 1-month lag; f, 2-month lag. The corresponding NDVI averages are from 50° to 80° N for Point Barrow, 10° to 30° N and 10° to 80° N for Mauna Loa, and 30° to 60° S for the South Pole.

to the photosynthetic capacity¹³ of the terrestrial surface in that zone and should be inversely related to atmospheric CO₂ concentrations. Figure 2 shows the latitudinal and temporal variation of the zonally averaged NDVI (weighted by land area) from April 1982 to October 1984. Consistent with considerations of growing seasons, the maximum zonally averaged and weighted NDVI occurs in summer at middle to high latitudes in the Northern Hemisphere. This pattern is a consequence both of the large proportion of land area and the highly seasonal behaviour of vegetation at these latitudes. The Southern Hemisphere south of 30° S has a low weighted NDVI because its vegetation is not as seasonal and the proportion of land area is less than in latitude zones north of 30° S. The weighted NDVI in equatorial areas is high throughout the year, and a desert belt is apparent as a depressed weighted NDVI from ~20° N to ~30° N. Figure 2 also shows differences in the weighted NDVI from 1982 to 1984. The 1982 global peak occurred in August, whereas the 1983 and 1984 global peaks occurred in July.

Results

The data presented in Figs 1 and 2 reveal an inverse relationship between the NDVI and the seasonal variations in atmospheric CO₂: in the Northern Hemisphere, the NDVI increased in spring to summer as the atmospheric CO₂ concentration decreased.

The NDVI data were plotted against the atmospheric CO₂ data from the Point Barrow, Mauna Loa and South Pole stations (Fig. 3). The South Pole data (Fig. 3c) showed no relationship between NDVI and CO₂: south of 30° S, the oceanic area is greater than land area which leads to smaller land/sea contrasts in air temperature. Consequently, seasonal behaviour of both the NDVI and the atmospheric CO₂ concentrations is less marked than in corresponding latitudes of the Northern Hemisphere (Figs 1, 2). Remote sources of CO₂ may also dominate or obscure local sources of CO₂ through atmospheric circulation and mixing in this region.

In contrast, the CO₂ data from Point Barrow and Mauna Loa and the NDVI data from 50 to 80° N and from 10 to 30° N, respectively (Figs 3a, b), showed a circular/elliptical relationship. At Point Barrow, the data showed no change in relative CO₂ concentration from April to June of 1982, 1983 and 1984, while the NDVI from 50 to 80° N increased (Fig. 3a). At Mauna Loa, however, the data showed a decrease in relative CO₂ concentration from May to June of 1982 and 1983, while the NDVI from 10 to 30° N again increased (Fig. 3b).

The Point Barrow results are consistent with heterotrophic respiration rates reported for soil and litter in the tundra and taiga zones between April and June^{36,37}. As soon as the ground thaws, soil and litter respiration releases a burst of CO₂ and this could explain the April to June displacement of the relative CO₂-NDVI relationship in Fig. 3a.

Atmospheric circulation and boundary-level mixing effects would be expected to delay equilibration of CO₂ concentrations by several weeks. A time lag between the NDVI and the measured atmospheric CO₂ concentrations would be expected because of the distance between the Point Barrow and Mauna Loa stations and mid-latitude forest sites which have maximum seasonality in photosynthesis. When the NDVI for month *n* was plotted against the atmospheric CO₂ concentration for month *n* + 1, approximately linear relationships were found with the Point Barrow and Mauna Loa data (Fig. 3d, e). The negative slopes in Fig. 3d and e indicate the increased absorption of CO₂ by increased amounts of terrestrial photosynthetic activity.

Inverse linear relationships, found between the Mauna Loa CO₂ concentrations, lagged by 2 months, and the NDVI for 10–80° N (Fig. 3f) showed that CO₂ oscillations at Mauna Loa represented the CO₂ response to the seasonal vegetation dynamics of the Northern Hemisphere and were not dominated by proximate sources and sinks. In addition, the globally averaged CO₂ cycle has a peak-to-peak amplitude of 4–5 p.p.m., about 75% of that at Mauna Loa, and is nearly in phase with

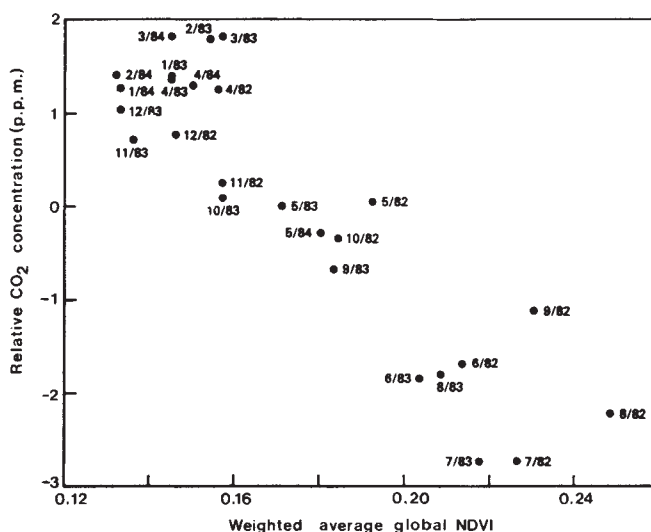


Fig. 4 The globally averaged atmospheric CO₂ concentration plotted against the globally averaged NDVI with a time lag of 1 month. The CO₂ data are from the global network of 20 NOAA/GMCC stations.

the cycle there. The annual cycle at Mauna Loa is thus characteristic of the globally averaged cycle.

Figure 4 shows the globally averaged CO₂ measurements from the 20 stations plotted against the globally averaged NDVI. When a lag of 1 month was introduced into the CO₂ data, an inverse linear relationship was found, suggesting that the global NDVI represents the photosynthetic capacity of the terrestrial biosphere.

Conclusions

Our data show the relationship between monthly variations in atmospheric CO₂ concentrations and terrestrial NDVI dynamics, which are, in turn, estimates of IPAR and photosynthetic capacity. Note that we have not considered factors such as respiration, decomposition, consumption of fossil fuels, oceanic processes and burning of vegetation, all of which influence atmospheric CO₂ concentrations. Our analysis demonstrates the measurable link between atmospheric CO₂ drawdown and terrestrial NDVI dynamics and suggests that there may be quantitative relationships between multi-temporal satellite data and atmospheric CO₂ drawdown. The inclusion of oceanic CO₂ uptake and release, respiration and decomposition processes, fossil fuel contributions and other CO₂ sources and sinks can be used to advance our quantitative understanding of the global CO₂ cycle.

We thank T. Goff, B. Rank, E. Matthews, K. Prentice and J. Lerner for assistance, T. J. Conway for the 1984 NOAA/GMCC flask data, and A. M. Heasty for comments. This research was supported in part by NASA Life Sciences Division, by NASA Earth Observation Division, and by NASA cooperative agreement NCC5-29 with Columbia University, as well as by the Carbon Dioxide Research Division, Office of Energy Research, US Department of Energy, under contract 19X-2229C with the Lamont-Doherty Geological Observatory of Columbia University with Martin Marietta Energy Systems, Inc., under contract DE-AC05-84OR21400.

Received 6 February; accepted 31 October 1985.

- Gates, D. M., Keegan, H. J., Schleter, J. C. & Weidner, V. P. *Appl. Opt.* **4**, 11–20 (1965).
- Knipling, E. B. *Remote Sens. Envir.* **1**, 115–119 (1970).
- Woolley, J. T. *Pl. Physiol.* **47**, 656–662 (1971).
- Kumar, R. & Silva, L. *Appl. Opt.* **12**, 2950–2954 (1973).
- Tucker, C. J. & Garratt, M. W. *Appl. Opt.* **16**, 635–642 (1977).
- Tucker, C. J. *Grass Forage Sci.* **35**, 177–182 (1980).
- Curran, P. J. *Phil. Trans. R. Soc. A* **309**, 257–270 (1983).

8. Kumar, M. & Monteith, J. L. in *Plants and the Daylight Spectrum* (ed. Smith, H.) 133-144 (Academic, London, 1982).
9. Daughtry, C. S. T., Galio, K. P. & Bauer, M. E. *Agron. J.* **75**, 527-531 (1983).
10. Hatfield, J. L., Asrar, G. & Kanemasu, E. T. *Remote Sens. Envir.* **14**, 65-76 (1984).
11. Wiegand, C. L. & Richardson, A. J. *Agron. J.* **76**, 543-548 (1984).
12. Asrar, G., Fuchs, M., Kanemasu, E. T. & Hatfield, J. L. *Agron. J.* **76**, 300-306 (1984).
13. Sellers, P. J. *Int. J. Remote Sens.* **6**, 1335-1372 (1985).
14. Keeling, C. D. *Proc. Carbon Dioxide Research Conference: Carbon Dioxide, Science and Consensus 2.1-2.62* (Department of Energy Conf.-82097, 1982).
15. Komhyr, W. D. *et al. J. geophys. Res.* **90**, 5567-5596 (1985).
16. Machta, L., Hansen, K. & Keeling, C. D. in *The Fate of Fossil Fuel CO₂ in the Oceans* (eds Anderson, N. R. & Malahoff, A.) 131-144 (Plenum, New York, 1977).
17. Woodwell, G. M., Houghton, R. A. & Tempel, N. R. *J. geophys. Res.* **78**, 932-940 (1973).
18. Fung, I. Y. in *Proc. 6th Oak Ridge natn. Lab. Life Sci. Symp. on the Global Carbon Cycle* (ed. Jacobs, V.) (US Department of Energy, Oak Ridge, in the press).
19. Heimann, M. M., Keeling, C. D. & Fung, I. Y. in *Proc. 6th Oak Ridge natn. Lab. Life Sci. Symp. on the Global Carbon Cycle* (ed. Jacobs, V.) (US Department of Energy, Oak Ridge, in the press).
20. Junge, C. E. & Czeplak, G. *Tellus* **20**, 422-434 (1968).
21. Machta, L. *Bull. Am. met. Soc.* **53**, 402-420 (1972).
22. Fung, I. Y., Prentice, K., Mathews, E., Lerner, J. & Russel, G. *J. geophys. Res.* **88**, 1281-1294 (1983).
23. Bolin, B. & Keeling, C. D. *J. geophys. Res.* **68**, 3899-3920 (1963).
24. Pearman, G. I. & Hyson, P. *J. geophys. Res.* **85**, 4457-4467 (1980).
25. Azevedo, A. E. thesis, Columbia University (1982).
26. Woodwell, G. M. *et al. Science* **222**, 1081-1086 (1984).
27. Woodwell, G. M. in *The Role of Terrestrial Vegetation in the Global Carbon Cycle: Measurement by Remote Sensing* (ed. Woodwell, G. M.) 1-17 (SCOPE 23, Wiley, New York, 1984).
28. Tucker, C. J., Townshend, J. R. G. & Goff, T. E. *Science* **227**, 369-375 (1985).
29. Justice, C. O., Townshend, J. R. G., Holben, B. N. & Tucker, C. J. *Int. J. Remote Sens.* **6**, 1278-1318 (1985).
30. Goward, S. N., Tucker, C. J. & Dye, D. G. *Vegetatio* (in the press).
31. Tucker, C. J., Vanpraet, C. L., Sharman, M. J. & Van Ittersum, G. *Remote Sens. Envir.* **17**, 233-249 (1985).
32. Tarpley, J. D., Schneider, S. R. & Money, R. L. *J. Climate appl. Met.* **23**, 491-493 (1984).
33. Holben, B. N. & Fraser, R. S. *Int. J. Remote Sens.* **5**, 145-160 (1984).
34. Schutz, C. & Gates, W. L. Preprint (Rand Corporation, R-1029-ARPA, Santa Monica, 1972).
35. Larcher, W. & Bauer, H. in *Physiological Plant Ecology* (eds Lange, O. L., Nobel, P. S., Osmond, C. B. & Ziegler, H.) 403-437 (Springer, Berlin, 1981).
36. Swift, M. J., Heal, O. W. & Anderson, J. M. *Decomposition in Terrestrial Ecosystems*, 308-309 (Blackwell, Oxford, 1979).
37. Linkins, A. E., Melillo, J. M. & Sinsabaugh, R. L. *Current Perspectives in Microbial Ecology* (eds Klug, M. J. & Reddy, C. A.) 572-579 (American Society of Microbiology, Washington, DC, 1984).

Solvation energy in protein folding and binding

David Eisenberg & Andrew D. McLachlan*

Molecular Biology Institute and Department of Chemistry and Biochemistry, University of California at Los Angeles, Los Angeles, California 90024, USA

* MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

We have developed a method for calculating the stability in water of protein structures, starting from their atomic coordinates. The contribution of each protein atom to the solvation free energy is estimated as the product of the accessibility of the atom to solvent and its atomic solvation parameter. Applications of the method include estimates of the relative stability of different protein conformations, estimates of the free energy of binding of ligands to proteins and atomic-level descriptions of hydrophobicity and amphiphilicity.

OF the forces that guide a polypeptide chain to its folded form in water, solvent interactions, including the hydrophobic interaction, are thought to be among the most important¹⁻⁴. Yet these interactions are currently evaluated by relatively primitive methods. This is in contrast to the other component energies that stabilize proteins, such as hydrogen bonds, for which atom-atom potential functions have been devised⁵⁻⁷. Present methods for estimating the contribution of solvation energy to protein stability include the assumption that the hydrophobic character of each amino-acid residue can be summarized by a single number, the amino-acid hydrophobicity⁸⁻¹². This is an oversimplification for amino acids such as Trp, Tyr, Glu, Gln, Lys and Arg which have both polar and apolar parts. Some authors have tried to overcome this limitation by devising more elaborate residue hydrophobicities to improve the description^{8,13,14}. Also an oversimplification is the commonly used approximation that the hydrophobic energy of a folded protein molecule is proportional to the total protein surface that is accessible to water, regardless of whether the exposed surface is apolar, polar or charged¹⁵.

To develop an explicitly atomic description of the interaction of water with a protein, we extend the ideas of Langmuir¹⁶, Cohn and Edsall¹⁷, and others^{18,19}. The basic assumption is that the free energy of interaction of a solute with water can be considered as a sum of energies of atomic groups. We follow Langmuir in using the exposed surface areas of a group as a measure of its interaction with solvent. In practice we use the solvent-accessible surface area of Lee and Richards²⁰ and other recent workers²¹⁻²⁴. This is defined as the area over which the centre of a water molecule of radius 1.4 Å can move while maintaining unobstructed contact with the group. In our method, the sign and strength of the water-solvent interaction are specified by the atomic solvation parameter (ASP) of each atom accessible to water. These values are determined, as described

below, from free energies of transfer⁸⁻¹². Thus, our method effectively combines two common approaches for evaluating hydrophobic forces: computation of solvent-accessible surface areas; and estimating energies from free energies of transfer. Our method, however, extends the first approach by weighting the effect of each atom by its polar or apolar character, and extends the second in permitting calculation of the solvation energy from the coordinates of individual atoms in each residue. It advances both methods in permitting estimates of the free energy of transfer of small molecules and of the contribution of solvation to the free energy of binding of small molecules to proteins.

Atomic contributions

Here we express the contribution of protein-solvent interactions to the free energy of protein folding as a sum over all atoms of the structure (except hydrogen atoms, which are not treated explicitly). The term for each atom is the product of its solvent-accessible area²⁰ and its ASP for transfer from the interior of a protein to aqueous solution. The transfer to solution of a single atom, *i*, immersed in protein is depicted in Fig. 1a. Let the accessible surface area of the atom be given by *A_i*, and its atomic solvation parameter be given by $\Delta\sigma_i$. Then the free energy of transfer (ΔG) is

$$\Delta G_i = \Delta\sigma_i A_i \quad (1)$$

Now consider the transfer to water of amino-acid residue *R* immersed in protein (Fig. 1a). We assume that the change in free energy for this process can be approximated as a sum of atomic terms, each like that of equation (2)

$$\Delta G_R = \sum_{\text{atoms } i} \Delta\sigma_i A_i \quad (2)$$

Note that the areas *A_i* now depend on the conformation of the amino acid, and that some atoms will be obscured by their

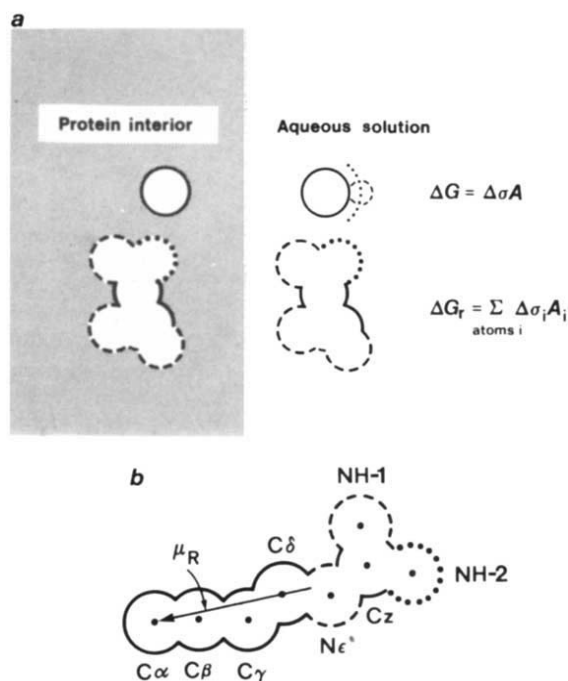


Fig. 1 *a*, Transfer of an atom and a molecule immersed in protein to water. The molecule contains three different types of atoms, represented by solid, dashed and dotted outlines. The definition of accessible surface area is illustrated for the atom at the right, where the centre of a sphere of radius 1.4 Å, tangent to the atom, sweeps out the accessible area. *b*, The atomic hydrophobic moment for the Arg side chain, shown as a vector running from the side-chain centre to the α -carbon atom. The chain takes the conformation specified in the CORELS⁴⁰ computer program dictionary. The dotted NH-2 atom is the most exposed nitrogen and is accordingly treated as the charged nitrogen.

Table 1 Amino-acid solvation free energies and group moments

Amino-acid residue	ΔG_{OBS}^*	$\Delta G_{\text{R}}^\dagger$	$\Delta G_{\text{TAN}}^\ddagger$	$\mu_a^\S$	$\cos \theta$
Gly	(0)	(0)	(0)	(0)	—
Ala	0.42	0.67	0.50	0	—
Val	1.66	1.5	1.50	0.48	0.84
Leu	2.32	1.9	1.80	1.0	0.89
Ile	2.46	1.9	—	1.2	0.99
Pro	0.98	1.2	—	0.18	0.22
Cys [¶]	1.34	0.38	—	0.17	0.76
Met	1.68	2.4	1.30	1.9	0.94
Thr	0.35	0.52	0.40	1.5	0.09
Ser	-0.05	0.01	-0.30	0.73	-0.67
Phe	2.44	2.3	2.50	1.1	0.92
Trp	3.07	2.6	3.40	1.6	0.67
Tyr	1.31	1.6	2.30	1.8	-0.93
Asn	-0.82	-0.60	—	1.3	-0.86
Gln	-0.30	-0.22	—	1.9	-1.0
Asp	-1.05	-1.2	—	1.9	-0.98
Glu	-0.87	-0.76	—	3.0	-0.89
His	0.18	0.64	0.50	0.99	-0.75
Lys	-1.35	-0.57	—	5.7	-0.99
Arg	-1.37	-2.1	—	10.0	-0.96
Group	$\Delta G_{\text{OBS}}^{\parallel}$ (ref. 17)	$\Delta G_{\text{R}}^\#$ (equation (3))			
-CH ₂ -	0.67	0.56			
-C ₆ H ₅	1.90	1.6			
-OH	-0.76	-0.66			
-CH ₂ -CONH-	-1.14	0.66			
Gly ionization	-3.7	-2.6			

Observed and calculated free energies of transfer are given in units of kcal mol⁻¹, calculated residue hydrophobic moments μ_a are given in kcal mol⁻¹ Å. ΔG_{OBS} and ΔG_{TAN} are experimental free energies of transfer relative to glycine. ΔG_{R} is the calculated value; μ_a is the calculated residue hydrophobic moment; and θ is the angle between the direction of the residue hydrophobic moment and a line from the α -carbon nucleus to the geometric centre of each side chain.

* ΔG_{OBS} is the observed value of π reported by Fauchere and Pliska²⁵ multiplied by 2.30 $RT = 1.36$ kcal mol⁻¹; π (side chain) is defined as $\log D(\text{acetyl amino-acid amide}) - \log D(\text{acetyl glycine amide})$, where D is the distribution coefficient for octanol/water.

† Calculated from equation (3).

‡ Observed values for transfer from ethanol to water from ref. 9.

§ Atom-based hydrophobic moment, defined in equation (5), below, using solvent-accessible surface areas from ref. 21 and atomic coordinates from ref. 39. These moments depend on the detailed conformation of the side chain, but not sensitively, because the corresponding values based on the coordinate dictionary supplied with the computer program CORELS⁴⁰ differ by an average of only 3%.

|| Experimental values from ref. 17 (from p. 212 except for the value for the -OH group which is from p. 206). Gly ionization is for the process: $\text{NH}_2\text{-CH}_2\text{-COOH} \rightarrow \text{NH}_3\text{-CH}_2\text{-COO}^-$.

Calculated from equation (3) as follows: -CH₂-, using the average accessible area of 8 -CH₂ groups in 5 residues; Gly ionization from tabulated areas²¹; and the -OH group as $\Delta G_{\text{R}}(\text{Ser}) - \Delta G_{\text{R}}(\text{Ala})$ and -C₆H₅- as $\Delta G_{\text{R}}(\text{Phe}) - \Delta G_{\text{R}}(\text{Ala})$, corresponding to the experimental values¹⁷.

¶ Both ΔG_{OBS} and ΔG_{R} refer to half-cystine. We were unable to derive a reliable value for -SH.

neighbours. We assume that conformation is not significantly changed during transfer.

The validity of equation (2) can be tested by a fit to experimental free energies of transfer. For proteins the fit is carried out by considering atoms in five classes: carbon, neutral oxygen and nitrogen (N/O), charged oxygen (O⁻), charged nitrogen (N⁺) and sulphur. Then the free energies of the common amino acids can be represented by equations of the type

$$\begin{aligned} \Delta G_{\text{R}} = & \Delta\sigma(\text{C}) \sum_{\text{C atoms } i} A(\text{C}_i, \text{R}) \\ & + \Delta\sigma(\text{N/O}) \sum_{\text{N, O atoms } i} A(\text{N/O}_i, \text{R}) \\ & + \Delta\sigma(\text{O}^-) \sum_{\text{O}^- \text{ atoms } i} A(\text{O}_i^-, \text{R}) \\ & + \Delta\sigma(\text{N}^+) \sum_{\text{N}^+ \text{ atoms } i} A(\text{N}_i^+, \text{R}) \\ & + \Delta\sigma(\text{S}) \sum_{\text{S atoms } i} A(\text{S}_i, \text{R}) \end{aligned} \quad (3)$$

in which $\Delta\sigma(\text{C})$ is the ASP for carbon and $A(\text{C}_i, \text{R})$ is the solvent-accessible surface area of carbon atom i in a standard conformation²¹ of residue R , and so forth for the other terms. The O⁻ term describes the charged oxygen atoms in Glu and Asp; the N⁺ term describes the charged nitrogen atoms in Lys, Arg and His residues; the N/O term describes all uncharged N and O atoms. In Asp, Glu and Arg the most exposed N or O atom is taken to be the atom on which the charge is concentrated. Only certain terms of equation (3) are needed to describe any given residue. For example, Asn requires only C and N/O terms. The pH dependence of charges is ignored, an assumption that might be thought to be poor for His, but the solvent-accessible areas of the N atoms are not large.

The assumption in equation (3) is that 20 residue hydrophobicities (ΔG_{R}) can be represented by five ASPs ($\Delta\sigma$ s), provided that suitable values of $\Delta\sigma$ can be found. We now show that a consistent set of $\Delta\sigma$ s can be derived from experimental data on free energies of transfer. The accessible areas A_i of residues in standard conformations can be calculated from known structures; we have adopted the areas calculated by Shrake and Rupley²¹, which are close to those found by C. Chothia (personal communication); they are for the amino acid X in a Gly-X-Gly sequence in a typical extended conformation. The values of the $\Delta\sigma$ s were determined by a linear least-squares fit of equations (3) to observed values for the free energy of

transfer, ΔG_R , of the amino-acid side chains. The measured ΔG s were those determined by Fauchere and Pliska²⁵ from transfer free energies of amino-acid residue analogues from *n*-octanol to water. Our estimated ASP values (with their standard deviations) are:

$$\begin{aligned}\Delta\sigma(C) &= 16 \pm 2 \text{ cal } \text{\AA}^{-2} \text{ mol}^{-1} \\ \Delta\sigma(N/O) &= -6 \pm 4 \\ \Delta\sigma(O^-) &= -24 \pm 10 \\ \Delta\sigma(N^+) &= -50 \pm 9 \\ \Delta\sigma(S) &= 21 \pm 10\end{aligned}$$

The quality of the fit is illustrated in Fig. 2, and the observed and calculated values are given in Table 1. The fit is generally good except for Lys and Cys. Note that the non-polar atoms C and S increase the free energy of the system as they are transferred from the interior of the protein to water. Polar atoms decrease the free energy in the same process; charged atoms cause a much larger decrease.

Solvation contribution

With $\Delta\sigma$ values, it is possible to estimate the solvation contribution to the free energy of protein folding, ΔG_s . This energy with respect to a reference state *r* is given by

$$\begin{aligned}\Delta G_s &= \Delta\sigma(C) \sum_{C \text{ atoms } i} (A_i - A_i^r) \\ &+ \Delta\sigma(N/O) \sum_{N, O \text{ atoms } i} (A_i - A_i^r) \\ &+ \Delta\sigma(O^-) \sum_{O^- \text{ atoms } i} (A_i - A_i^r) \\ &+ \Delta\sigma(N^+) \sum_{N^+ \text{ atoms } i} (A_i - A_i^r) \\ &+ \Delta\sigma(S) \sum_{S \text{ atoms } i} (A_i - A_i^r)\end{aligned}\quad (4)$$

in which *A* is the solvent-accessible surface area of an atom in the folded state and *A*^r is that in the reference state²¹. In the usual way in which equation (4) is used, the reference state does not matter because the energies of the same polypeptide chain are compared in two conformations having a common reference state. This is true for the application described below.

Improperly folded structures

One potential use of equation (4) is to assess the stabilities of modified or redesigned proteins. To illustrate such a calculation, the models of Novotný *et al.*¹ for incorrectly folded proteins are useful. These models were generated by exchanging the 113 side chains on the correctly folded α -carbon backbone of haemerythrin²⁶ from *T. dyscritum* with those from the correctly folded, β -sheet backbone of an immunoglobulin variable light-chain (*V_L*) domain²⁷ having the same number of residues. Novotný *et al.* used an energy-minimizing computer program to make small annealing adjustments in coordinates, and to reduce the potential energies of the two incorrectly folded structures to local minima. These annealed structures then had satisfactory van der Waals contacts, with normal valence angles and bond lengths, but no explicit account had been taken of solvation effects. Novotný *et al.*¹ noted that there is greater exposure of apolar residues in the misfolded structures. To these model misfolded structures, as well as to their energy-annealed, correctly folded analogues, we have applied equation (4) to estimate the solvation contribution to the free energy of folding. Our procedure consists simply of: first, computing²² solvent-accessible atomic areas of the four structures from atomic coordinates; and second, weighting each atomic area by its ASP, and summing.

The solvation free energy is lower for the correct structures (Table 2). Both real structures have solvation free energies lower

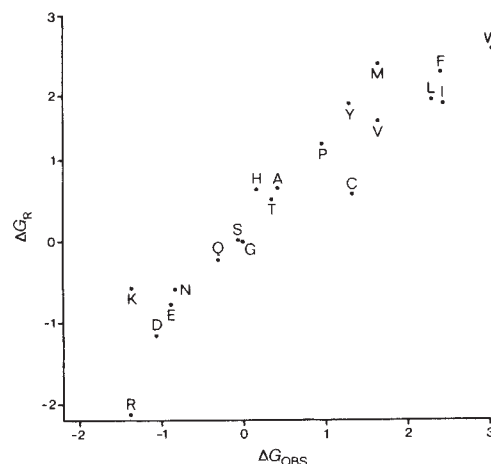


Fig. 2 Comparison of experimental values ΔG_{OBS} for free energies of transfer for amino-acid residues with the values of ΔG_R calculated from equation (3), using ASPs determined by linear least squares. The details of the least-squares fit are as follows: Fauchere and Pliska²⁵ deduced the hydrophobic parameter π for 19 amino-acid side chains as the difference between the logarithm of the measured distribution coefficients of the acetyl amino-acid amides and acetyl-glycine amide. Equation (3) was fitted to the free energies of transfer, $2.303 RT\pi = \Delta G_{OBS}$, with the aid of NAG library routines G02BFF and G02CHF (Numerical Algorithms Group Ltd, Mayfield House, 256 Banbury Road, Oxford OX2 7DE, UK). In this fitting procedure, the difference between ΔG_R and ΔG_{OBS} is minimized by least-squares determination of the ASP values. The values of *A* for the appropriate atoms in each side chain are taken from ref. 21. For all residues, only the differences in area with the Gly residue are considered. Thus backbone atoms are ignored, except for the difference in exposed area for the α -carbon atom of glycine and of the other side chains. The deviation of points of the figure from a straight line can be explained by any of: (1) errors in measurements of ΔG_{OBS} ; (2) inappropriate areas in equation (3); or (3) faulty assumptions in the formulation of equation (3) (see below). The multiple correlation coefficient *R* is 0.95, and the value of its square corrected for the number of parameters (ASPs), R^2_{COR} , is 0.88. The R^2_{COR} suggests that the separate ASPs for O^- and N^+ are justified, as is the sulphur ASP: with only a single type of charged atom, R^2_{COR} falls to 0.85 and *R* to 0.93; for only C/S and N/O ASPs, R^2_{COR} falls to 0.11 and *R* to 0.45. Fits to other hydrophobicity scales give different values for the ASPs. For example, the consensus scale of Eisenberg *et al.*¹², an average of five scales, yields *R* = 0.97 for a five-term fit. The ASP values for the consensus fit are similar to those reported here, but the O^- and N^+ ASPs have larger values. Despite the better fit to the consensus scale, the ASPs based on the data of Fauchere and Pliska²⁵ have been used in our calculations, in part because the absolute level of the consensus scale is unknown and in part because the new ASPs for the charged atoms seem more reasonable.

than the hypothetical, unfolded reference state by some 100 kcal mol⁻¹, ~1 kcal per residue. Moreover, the solvation free energies for the real structures are markedly lower than for the incorrectly folded ones. Not only does each correctly folded structure have lower solvation free energy than the same polypeptide sequence folded improperly, but also each correct structure has lower free energy than the other sequence folded with the same backbone pattern. These results can be contrasted with the behaviour of traditional energy terms¹, which are not significantly different in the folded and misfolded structures.

When contributions to the solvation energy of folding for the natural structures are examined, interesting details emerge. For example, in the haemerythrin structure, 95 of the 113 side chains stabilize the solvation energy. Only one side chain, Asp 11, is destabilizing by so much as 1 kcal mol⁻¹. For this side chain, both oxygen atoms are somewhat shielded from solvent relative to the extended reference structure. But for the protein as a whole, the process of folding from the extended reference state

Table 2 Solvation free energies of folding ΔG_f for an immunoglobulin V_L domain and haemerythrin

	V_L immunoglobulin domain	Haemerythrin
Correct fold	-106	-113
Incorrect fold (with the same sequence)	-72	-96
Net stabilization of correct structure	-34	-17

Values are calculated for the natural structures and the incorrectly folded structures (with exchanged amino-acid sequences) constructed by Novotný *et al.* Energies are given in kcal mol⁻¹. Equation (4) is used, with areas determined by the program ACCESS²². The uncertainty in the net stabilization of the correct structure was estimated from the standard deviations in the ASPs and by assuming that the fractional uncertainty in the sums of the atomic areas of each type is 0.05; this results in an overall uncertainty of ~10 kcal mol⁻¹.

yields -159 kcal for sequestering the 606 C and 3 S atoms away from the solvent (the sum of the first and last terms of equation (4)), while the partial covering of the 297 N/O, 17 O⁻, and 21 N⁺ atoms consumed only 25, 12 and 11 kcal mol⁻¹, respectively. Significant contributions to the stability of the folded state come from the apolar carbon atoms of the 29 Lys, Arg, Asp and Glu residues (a total of 41 kcal mol⁻¹). These contributions would be obscured or neglected in energetic models that are cast in terms of whole residue hydrophobicities; clearly, whole-residue hydrophobicities cannot account adequately for the folding energies of proteins.

Agreement with experimental data

Although it is reassuring that the estimated free energy of folding is lower for the real protein structures than for the incorrectly folded chains, this result falls well short of experimental verification of the method proposed here. Here we examine its assumptions and compare its predictions with experiment.

The principal assumption in the use of equation (3) to derive the ASPs is that the free energy of transfer is a linear function of the accessible areas of the constituent atoms of the transferred molecule. This assumption is supported indirectly by the observed linearity of free energy with surface area for the transfer of many compounds from various non-aqueous liquids into water^{15,28-31}. Of more direct bearing on the assumption is the accumulated evidence that molecular free energies of transfer can often be reckoned as a sum of energies for atomic groups¹⁹. This is not a new idea. Cohn and Edsall¹⁷ made a critical analysis of evidence which suggested that the free energy of transfer of a chemical group such as -CH₂-, from ethanol and other apolar solvents to water is often independent of the molecule in which it is found. Some values for the free energy of transfer for various groups as compiled by Cohn and Edsall are given in Table 1, where they are compared with our new values calculated from equation (3). Generally, the computed values agree well in sign and magnitude with the experimental values, especially considering the diversity of the groups considered. The calculated values for side-chain transfers are also in reasonable agreement with the well-known measured values of Nozaki and Tanford⁹ (Table 1). An exception to agreement in Table 1 is the free energy of transfer of the glycyl residue, for which equation (3) gives the wrong sign. This is probably because our model takes virtually no account of the partial charges on the amide N and O atoms. The reason is that the backbone atoms are omitted in the fitting procedure used to determine the ASPs, because the data being fitted are all differences between acetyl glycyl amide and the corresponding derivative of other amino acids. From the value in Table 2 for the free energy of transfer of the glycyl residue from ethanol to water, it is possible to estimate a crude value for the ASP for the amide N and O atoms at ~-50 cal

Å⁻² mol⁻¹. If this value is qualitatively correct, then our solvation energies of folding are somewhat overestimated. To resolve this uncertainty, it will be necessary to fit a larger database of comparable free energies of transfer, which is not yet available.

An additional assumption in the use of equation (4) is that a non-polar liquid is an adequate model for the interior of a protein, as the ASPs derived from transfers between liquids are applied by equation (4) to transfer from liquid to protein interior. This assumption has been examined by Richards²⁴, who also considered amino-acid crystals as models for the protein interior. Richards concluded that an ASP somewhere in the range of 13-26 cal Å⁻² mol⁻¹ is appropriate for the transfer of apolar atoms. Our $\Delta\sigma(C)$ and $\Delta\sigma(S)$ values of 16 and 21 cal Å⁻² mol⁻¹ lie in this range, but may deserve revision as further free-energy data become available. The value for $\Delta\sigma(S)$, resting on only two measurements, is particularly subject to revision.

Atomic hydrophobic moments

In earlier work we described the amphiphilicity (asymmetry of hydrophobicity) of segments of regular protein secondary structure in terms of hydrophobic moments³². It was found that the hydrophobic dipole moments of neighbouring segments of secondary structure tend to oppose each other in correctly folded proteins¹², but not in incorrectly folded ones³³, and that hydrophobic moments can be used to classify helices on the basis of amino-acid sequences^{34,35} and to detect periodicities^{36,37}. These applications of hydrophobic moments were all formulated in terms of residue moments and can perhaps be made more precise with atomic-level moments.

The hydrophobic moment can be defined as an atomic property using the ASP values determined here. The hydrophobic dipole moment of a group of atoms is given by a sum over atoms i

$$\mu_a = \sum_{\text{atoms } i} \Delta\sigma_i A_i \mathbf{r}_i - \langle \Delta\sigma A \rangle \sum_{\text{atoms } i} \mathbf{r}_i \quad (5)$$

in which $\mathbf{r}_i$ is a vector from any origin to the position of atom i , and where the brackets indicate the mean value for all atoms of the group. The second term ensures that the value of μ_a is independent of the choice of origin. When the sum in equation (5) is restricted to the atoms in a single side chain, a residue hydrophobic dipole moment is defined. This quantity is a measure of the amphiphilicity of the side chain, as distinct from its hydrophobicity. The values of these moments are tabulated in Table 1.

From these values it is apparent that the residues of greatest intrinsic amphiphilicity are Arg, Lys and Glu. In contrast, the most hydrophobic residues, Trp, Phe, Leu and Ile, all have small amphiphilicities. The direction of the hydrophobic moment is also defined by equation (5). This is expressed in Table 1 as the cosine of the angle between the direction of the moment and the direction of the vector from the α carbon to the side-chain centre. For the highly amphiphilic residues, the direction of the moment is nearly antiparallel to the α -carbon side-chain vector ($\cos \theta$ near to -1). This is illustrated in Fig. 1b for Arg. Some of the most hydrophobic side chains (for example, Ile, Met and Phe) have moments nearly parallel to the α -carbon-centre vector ($\cos \theta$ near to +1). Note that when the phenolic oxygen (with negative $\Delta\sigma$) is added to Phe to form Tyr, it reverses the direction of the moment (from +0.92 to -0.93).

Implications and extensions

How does the solvation free energy defined in equation (4) correspond to the component of the free energy of protein folding first described by Kauzmann as hydrophobic³? Kauzmann suggested that exposed apolar side chains increase the free energy of the system because they decrease the entropy of water. This effect on the free energy is described in equation (4) by the terms $\Delta\sigma(C)$ and $\Delta\sigma(S)$. Kauzmann's approach is extended in equation (4) by including the terms $\Delta\sigma(N/O)$,

$\Delta\sigma(\text{O}^-)$ and $\Delta\sigma(\text{N}^+)$. This has been done to introduce information from empirical free-energy measurements in a self-consistent way, and to impose a penalty in free energy for burial of charged and polar groups with a magnitude determined from experiments. These terms probably include contributions from electrostatic³⁸ and hydrogen-bonding effects as well. A full understanding of these contributions would require a knowledge of the structure of the molecules used to derive the ASPs, including states of protonation and hydration in both solvents. In short, ΔG_s includes contributions to the free energy of folding that go beyond the entropy change of the solvent. For this reason we call this energy the solvation free energy, rather than the hydrophobic free energy.

One use of equation (4) is in assessing the stability of proposed new proteins. These can either be genetically redesigned proteins or proteins designed for peptide synthesis. The energy of any model described in atomic coordinates can be tested against equation (4), although it may be necessary to supplement this equation with terms for other components of the folding energy. Another use is in energy-coupled X-ray phase refinement of protein structures³⁹⁻⁴². In existing procedures, a model is determined that is the best compromise between the potential energy

of the structure and the fit to the X-ray data. These methods do not go far enough in that it is the free energy, including solvation effects, that determines the structure. These effects could be incorporated by using ASPs as suggested here, along with the analytical method of Richmond²³, for computing accessible surface areas and their spatial derivatives. Similarly, the addition of a solvation energy term of the form of equation (4) could be made to molecular dynamics calculations. Finally, we note that the general method of computing solvation energy expressed in equation (4) could be developed further to compute the solvation contribution to the free energy for association of proteins with membranes or nucleic acids.

We thank Drs Novotný, Brucoleri and Karplus for making available the coordinates of the incorrectly and correctly folded protein structures described in ref. 1; Dr C. Chothia for his values of accessible surface areas of amino acids in extended polypeptide chains; Dr Richmond for use of his computer program ACCESS; and Drs D. Abrahams, F. R. Salemme and C. Schutt for discussions. D.E. acknowledges a Guggenheim fellowship and support from NIH and NSF. A computer program to calculate solvation energy according to equation (4) is available from the authors.

Received 19 July; accepted 30 October 1985.

- Novotný, J., Brucoleri, R. & Karplus, M. *J. molec. Biol.* **177**, 787-818 (1984).
- Kauzmann, W. *Adv. Protein Chem.* **14**, 1-63 (1959).
- Creighton, T. E. *Proteins* (Freeman, San Francisco, 1984).
- Tanford, C. *The Hydrophobic Effect* 2nd edn (Wiley, New York, 1980).
- Karplus, M. & McCammon, A. A. *Rev. Biochem.* **52**, 263-300 (1983).
- Nemethy, G. & Scheraga, H. A. *Q. Rev. Biophys.* **10**, 239-352 (1977).
- Levitt, M. & Warshel, A. *Nature* **253**, 694-698 (1975).
- Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H. & Zehfus, M. H. *Science* **229**, 834-838 (1985).
- Nozaki, Y. & Tanford, C. *J. biol. Chem.* **246**, 2211-2217 (1971).
- Janin, J. *Nature* **277**, 491-492 (1979).
- Wolfenden, R., Anderson, L., Cullis, P. M. & Southgate, C. C. B. *Biochemistry* **20**, 849-855 (1981).
- Eisenberg, D., Weiss, R. M., Terwilliger, T. C. & Wilcox, W. *Faraday Symp. chem. Soc.* **17**, 109-120 (1982).
- Guy, H. R. *Biophys. J.* **47**, 61-70 (1985).
- Frommel, C. *J. theor. Biol.* **111**, 247-260 (1984).
- Chothia, C. *Nature* **248**, 338-339 (1974).
- Langmuir, I. *Colloid Symp. Monogr.* **3**, 48-75 (1925).
- Cohn, E. J. & Edsall, J. T. in *Proteins, Amino Acids and Peptides as Ions and Dipolar Ions*, Ch. 9 (ACS Monograph Series, Reinhold, New York, 1943).
- Tanford, C. *J. Am. chem. Soc.* **84**, 4240-4247 (1962).
- Hansch, C. & Leo, A. in *Substitution Constants for Correlation Analysis in Chemistry and Biology* (Wiley, New York, 1979).

- Lee, B. & Richards, F. M. *J. molec. Biol.* **55**, 379-400 (1971).
- Shrake, A. & Rupley, J. A. *J. molec. Biol.* **79**, 351-371 (1973).
- Richmond, T. J. & Richards, F. M. *J. molec. Biol.* **119**, 537-555 (1978).
- Richmond, T. J. *J. molec. Biol.* **178**, 63-89 (1984).
- Richards, F. M. A. *Rev. Biophys. Bioengng* **6**, 151-176 (1977).
- Fauchere, J.-L. & Pliska, V. *Eur. J. med. Chem.-Chim. ther.* **18**, 369-375 (1983).
- Stenkamp, R. E., Sieker, L. C. & Jensen, L. H. *Acta crystallogr.* **B38**, 784-792 (1978).
- Segal, D. M. et al. *Proc. natn. Acad. Sci. U.S.A.* **71**, 4298-4302 (1974).
- Hermann, R. B. *J. phys. Chem.* **76**, 2754-2759 (1972).
- Nandi, P. K. *Int. J. Peptide Protein Res.* **8**, 253-264 (1976).
- Gelles, J. & Klapper, M. H. *Biochim. Biophys. Acta* **533**, 465-477 (1978).
- Creighton, T. E. *J. molec. Biol.* **129**, 235-264 (1979).
- Eisenberg, D., Weiss, R. M. & Terwilliger, T. C. *Nature* **299**, 371-374 (1982).
- Eisenberg, D., Wilcox, W. & McLachlan, A. D. *Proc. UCLA Symp. Protein Structure, Folding and Design* (in the press).
- Eisenberg, D., Schwarz, E., Komaromy, M. & Wall, R. *J. molec. Biol.* **179**, 125-142 (1984).
- Pownall, H. J., Knapp, R. D., Gatto, A. M. Jr & Massey, J. B. *FEBS Lett.* **159**, 17-23 (1983).
- Eisenberg, D., Weiss, R. M. & Terwilliger, T. C. *Proc. natn. Acad. Sci. U.S.A.* **81**, 140-144 (1984).
- Finer-Moore, J. & Stroud, R. M. *Proc. natn. Acad. Sci. U.S.A.* **81**, 155-159 (1984).
- Warshel, A. & Russell, S. T. *Q. Rev. Biophys.* **17**, 283-422 (1984).
- Diamond, R. *J. molec. Biol.* **82**, 371-391 (1974).
- Susman, J. L., Holbrook, S. R., Church, G. M. & Kim, S.-H. *Acta crystallogr.* **A33**, 800-804 (1977).
- Levitt, M. *J. molec. Biol.* **82**, 393-420 (1974).
- Konnert, J. H. *Acta crystallogr.* **A32**, 614-617 (1976).

LETTERS TO NATURE

Inadequacy of Coulomb's friction law for particle assemblies

K. Kendall

ICI New Science Group, PO Box 11, The Heath, Runcorn, Cheshire WA7 4QE, UK

Wherever powders are manipulated by mechanical means, adhesive and frictional forces act between the solid grains. For two centuries, these forces have been described by Coulomb's law¹, a law which originally proved useful in designing civil engineering structures, and which has since been profitably extended to the understanding of the powder flows in chemical plants². However, we have now found that Coulomb's law is not adequate to explain the compaction of fine powders in the manufacture of ceramic articles, where it has been observed that friction appears to increase for smaller particles. This apparent increase in friction is explained by considering in detail the contact of loaded, adhesive, elastic spheres.

Consider the ideal experiment (Fig. 1a) in which a bed of particles is loaded onto and sheared along a polished plate of the same material³. The nominal shear stress F/A (where F is

force and A is area) required to slide the powder across the plate under the normal stress W/A (where W is normal load) has usually been interpreted by Coulomb's law:

$$F/A = K + \mu W/A \quad (1)$$

where K represents the cohesive shear strength of the particle/plate assembly (that is, the shear stress experienced at zero normal load) and μ is the coefficient of friction between the surfaces⁴. This law, in combination with the geometrical constraints on powder motion first mentioned by Reynolds in 1885, has been much used to describe the deformation of powders and soils in complex stress states⁵⁻⁷. Here we demonstrate that equation (1) does not follow the experimental results satisfactorily, and we propose a new argument which provides an improved fit to the observed behaviour.

The problem of applying Coulomb's law to assemblies of fine particles is that the friction coefficient calculated from equation (1) does not generally match that measured in a standard friction test between macroscopic blocks of material. First, the calculated μ is often higher than expected: for example, sand particles of diameter 0.5 mm in water give $\mu = 0.45$ whereas two large quartz blocks give $\mu = 0.3$ (Fig. 1b). A second difficulty is that the friction coefficient increases as the grains are made smaller,

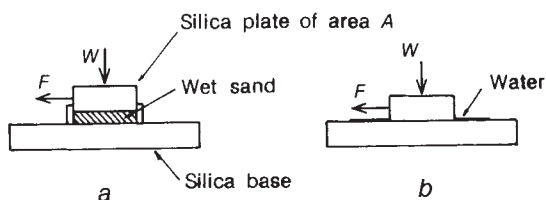


Fig. 1 *a*, Experimental set-up for measuring friction of wet sand particles pressed against a smooth silica plate. *b*, Friction measurement on macroscopic silica plates.

rising to $\mu = 0.6$ for wet sand of 0.01 mm diameter. Finally, the friction of powders does not remain constant as the normal pressure is increased, but falls substantially at high loads, in conflict with Coulomb's law. Therefore, a new theory is required to account for these curious effects.

The fundamental origin of Coulomb's law may be seen by considering a spherical particle in frictional contact with a smooth plane under normal load w and shear force f (Fig. 2). Because the particle experiences a surface attraction towards the plane as a result of van der Waals' force w_s , the total load on the contact is $w_s + w$ and the frictional force is thus

$$f = \mu(w_s + w) = k + \mu w \quad (2)$$

which is the form of equation (1). Thus, Coulomb's law is the simplest expression for frictional resistance, taking into account both the external load and the interfacial attraction between the particle and the plane.

Equation (2) has been predominant in describing soil behaviour⁸; it has been re-derived by Derjaguin⁹, and verified by Lazarev¹⁰ for flat surfaces in extended molecular contact. It is used to this day to account for the friction of powders^{11,12} and fibres¹³ although it does not follow the results satisfactorily.

The difficulty with the simple Coulomb model is the assumption that the external load and the interfacial force are independent. This assumption is wrong, as indicated by recent studies on the contact of elastic spheres exhibiting surface attractions¹⁴. For a sphere of diameter D resting on a plane surface of the same material with interface energy Γ (the energy required to break unit area of contact) and normal load w , the effective load pressing the sphere onto the plane has been shown to be¹⁴

$$\frac{3\pi D\Gamma}{2} + w + \left[3\pi D w \Gamma + \left(\frac{3\pi D\Gamma}{2} \right)^2 \right]^{1/2} \quad (3)$$

so that the equation for shear force becomes

$$f = \frac{\mu 3\pi D\Gamma}{2} + \mu w + \mu \left[3\pi D w \Gamma + \left(\frac{3\pi D\Gamma}{2} \right)^2 \right]^{1/2} \quad (4)$$

This equation is similar to the Coulomb form of equation (2) in that the first term stems from surface attraction and the second term is due to the external load; it differs, however, from the Coulomb model in the extra term which gives the interdependence of the molecular forces and the applied load. It also differs from the recent theory for the sliding of adhesive spheres suggested by Savkoor and Briggs¹⁵.

Equation (4) is useful in describing some of the anomalies reported previously. For example, it explains the nonlinearity of plots of friction force against normal load noted by Briscoe and Kremnitzer on polyethylene terephthalate fibres¹³. Figure 2 shows how the new theory follows the experimental data. The spread of friction results suggested a value of interface energy Γ of ~ 0.01 – 0.04 J m^{-2} , in fair agreement with the value calculated from direct adhesion tests. Note that the apparent friction coefficient (the slope of the curve) is twice the true μ near zero load, but decreases to the correct value at high loads, or for large particles.

The predictions of equation (4) were tested experimentally, using the apparatus of Fig. 1*a* to determine the friction of wet

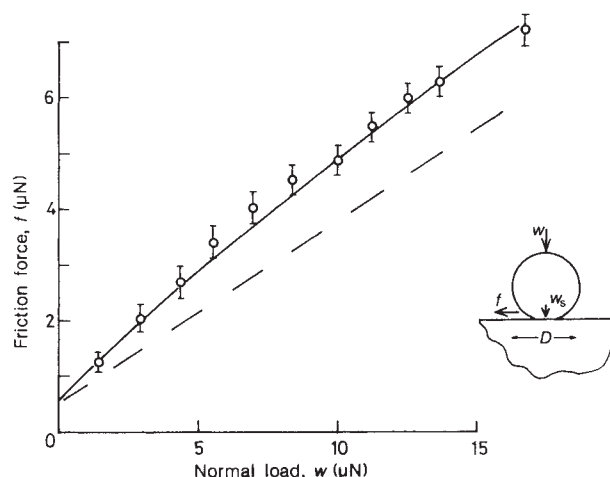


Fig. 2 Contact between a particle and a plate exhibiting interfacial attractive energy Γ under a normal load w and shear force f . The surface attraction between the spheres leads to a nonlinear relation between f and w [equation (4)] which differs from Coulomb's law at low loads but approaches Coulomb behaviour at high loads. —, Curve derived from equation (4) with $\mu = 0.33$, $\Gamma = 0.01 \text{ J m}^{-2}$; ---, curve derived from Coulomb's law, $\mu = 0.33$; $\circ$, results from ref. 13.

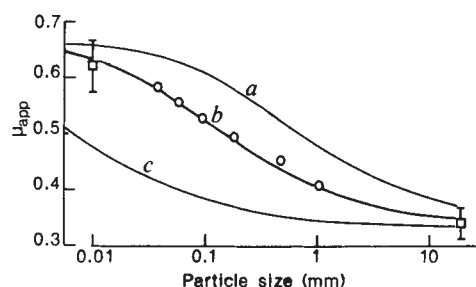


Fig. 3 Plot of μ_{app} against particle size derived from equation (5) for comparison with experimental results on wet sand. A good fit with the experimental results was obtained when the true friction coefficient was assumed to be 0.332 and the interface energy 2 J m^{-2} . *a*–*c*, Curves derived from equation (5) with $\Gamma = 10, 2$ and 0.1 J m^{-2} , respectively. $\square$, Our results; $\circ$, those of Rowe³.

silica powder. Sand was sieved into fractions and was loaded between smooth silica sheets, then sheared to measure the friction force. This friction force remained constant over many traversals of the silica, suggesting that the deformation was largely elastic, as assumed in the theory. Various loads were used, giving normal pressures close to 100 kPa, and the slope of the friction force/normal force curve was calculated. This slope was the apparent friction coefficient, μ_{app} , which was then compared with the theoretical value computed by differentiation of equation (4):

$$\mu_{app} = \frac{df}{dw} = \mu \left[1 + \left(1 + \frac{4PD}{3\pi\Gamma} \right)^{-1/2} \right] \quad (5)$$

where $P \approx w/D^2$ is the normal applied pressure.

Figure 3 shows the experimental results for μ_{app} as a function of the particle size at a normal pressure of 100 kPa. The apparent friction coefficient achieved a maximum of 0.63 for the smallest grains, falling steadily for larger particles, and reaching a value of 0.34 for the macroscopic plates. The theoretical curve plotted from equation (5) gave a remarkably good fit to the data when the interface energy Γ was taken to be 2 J m^{-2} and the true friction coefficient was $\mu = 0.332$. Although this value of interface energy seems somewhat high for wet silica, a similar discrepancy has been noted in electron microscope studies of

adhering metal particles, probably owing to the nonspherical nature of the grains¹⁶.

It seems likely, therefore, that the new theory presented here offers the correct explanation for the failure of Coulomb's law in describing the odd frictional behaviour of fine powders, particularly those materials of high surface area which are used in the field of modern ceramics.

Received 13 August; accepted 31 October 1985.

1. Coulomb, C. A. *Mém Acad. Sci. Savants Étrangers* 7, 343-382 (1773).
2. *Proc. IUTAM Symp. on Deformation and Failure of Granular Materials*, (Balkema, Delft, 1982).
3. Rowe, P. W. *Proc. R. Soc. A* 269, 500-527 (1962).
4. Bowden, F. P. & Tabor, D. *Friction and Lubrication of Solids* Pts I and II (Clarendon, Oxford, 1950 and 1964).
5. Reynolds, O. *Phil. Mag.* 20, 468-481 (1885).
6. Bolton, M. A *guide to Soil Mechanics* Ch. 4 (Macmillan, London, 1979).
7. Thornton C. & Barnes, D. J. *Proc. IUTAM Symp. on Deformation and Failure of Granular Materials*, 69-77 (Balkema, Delft, 1982).
8. Terzaghi, K. *Theoretical Soil Mechanics*, 7, (Wiley, New York, 1943).
9. Derjaguin, B. V. *Phys. Khim.* 5, 1163 (1934).
10. Derjaguin, B. V. & Lazarev, V. P. *Kolloid. Zh.* 1, 293 (1935).
11. Chappuis, J., Hoepfner, D. W. & Neumann, A. W. *Tribology Ser.* 7, 71-80 (Elsevier, Oxford, 1982).
12. Omenyi, S. N., Chappuis, J. & Neumann, A. W. *J. Adhesion* 13, 131-139 (1981).
13. B. J. Briscoe, & Kremnitzer, S. L. *J. Phys. D* 12, 505-516 (1979).
14. Johnson, K. L., Kendall, K. & Roberts, A. D. *Proc. R. Soc. A* 324, 301-313 (1971).
15. Savkoor, A. R. & Briggs, G. A. D. *Proc. R. Soc. A* 356, 103-114 (1977).
16. Tholen, A. R. *Tribology Ser.* 7 263-277 (Elsevier, Oxford, 1982).

Secondary antiproton production in relativistic plasmas

C. D. Dermer & R. Ramaty

Laboratory for High Energy Astrophysics, NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

Recent measurements¹⁻³ of the low-energy flux of antiprotons ($\bar{p}$) in cosmic rays give ratios of $\bar{p}/p$ which are greater by a factor of 3-5 than predicted by the simple 'leaky box' model of cosmic-ray propagation. Here we report on an investigation of the possibility that the low-energy $\bar{p}$ excess reflects an origin in p-p collisions in relativistic plasmas. Because of both target and projectile motion in such plasmas, the antiproton production threshold in the frame of the plasma is much lower than the threshold of antiproton production in cosmic ray interactions with ambient matter. The spectrum of the resultant antiprotons therefore extends to much lower energy than in the cosmic ray case. Sites at which matter is accreted by compact objects could be sources of the low energy antiprotons. Escape of the antiprotons could be mediated by anti-neutron production.

Golden *et al.*¹ have reported an antiproton to proton ($\bar{p}/p$) ratio of $5.2(\pm 1.5) \times 10^{-4}$ in the cosmic radiation in the energy range 4.7-11.6 GeV. Bogomolov *et al.*², on the basis of only two events, report a $\bar{p}/p$ ratio of $6(\pm 4) \times 10^{-4}$ at energies between 2 and 5 GeV. Buffington *et al.*³ report a $\bar{p}/p$ ratio of $2.2(\pm 0.6) \times 10^{-4}$ between 130 and 320 MeV, and also determine that the antihelium to helium ratio $\bar{\alpha}/\alpha < 2.2 \times 10^{-5}$ in the energy range 130-370 MeV per nucleon. Upper limits on previous antinucleus searches are also given here.

The simple 'leaky box' model of cosmic-ray propagation predicts a $\bar{p}/p$ ratio smaller by a factor of 3-5 than the values measured in the Golden and Bogomolov experiments. Because of the kinematic cutoff associated with the high $\bar{p}$ production threshold in p-p collisions when one of the protons is at rest, the $\bar{p}/p$ ratio predicted by this model is about two orders of magnitude lower than the value reported by Buffington *et al.*, even after the effects of solar modulation are taken into account⁴. The low $\bar{\alpha}/\alpha$ ratio, in comparison with the $\bar{p}/p$ ratios, suggests a secondary origin of the $\bar{p}$, although primary cosmic-ray antiproton theories, with subsequent breakdown of the $A > 1$ antinuclei, have been proposed (see ref. 5).

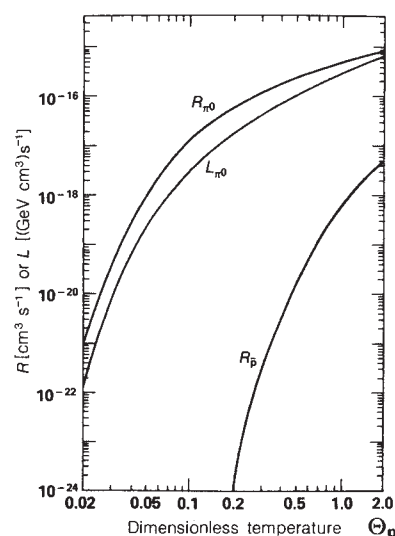


Fig. 1 The rate coefficients for secondary π^0 and $\bar{p}$ production from p-p collisions in a relativistic plasma at temperature $\theta_p = kT/m_p c^2$. Also shown is the π^0 luminosity coefficient giving the total energy in secondary π^0 γ -rays produced at temperature θ_p .

Theories of secondary $\bar{p}$ production in p-p collisions in the Galaxy are constrained by the observed galactic γ -ray luminosity. The observed antiproton flux $\phi_p(E)$ implies a total galactic $\bar{p}$ production rate

$$Q(\bar{p}) \sim \frac{4\pi V}{c} \int_0^\infty dE \frac{\phi_p(E)}{\beta\tau(E)} \sim (3.5 \times 10^{-14} V/\tau) \bar{p} \text{ s}^{-1}$$

where V is the galactic confinement volume and τ is the average galactic $\bar{p}$ residence time. Taking $V \sim \pi(15 \text{ kpc})^2 (1 \text{ kpc}) \sim 1.2 \times 10^{67} \text{ cm}^3$ and $\tau \sim 10^7 \text{ yr}$ gives $Q(\bar{p}) \sim 2.1 \times 10^{39} \bar{p} \text{ s}^{-1}$. If the production of a secondary $\bar{p}$ is accompanied by the production of $\langle n \rangle$ π^0 -decay γ -rays, the resultant γ -ray luminosity of the galaxy $Q(\gamma) \sim \langle n \rangle Q(\bar{p})$. In the case of the production of > 100 MeV photons in cosmic-ray interactions, $\langle n \rangle_{\text{cr}} \sim 2.7 \times 10^3$, from the calculations of Stephens and Badhwar⁶ and Tan and Ng⁴. If the $\bar{p}$ in the cosmic radiation are entirely cosmic-ray secondaries, the > 100 MeV luminosity of the Galaxy should, therefore, be at least $5.7 \times 10^{42} \gamma \text{ s}^{-1}$, in contrast to the value of $\sim 2.5 \times 10^{42} \gamma \text{ s}^{-1}$ deduced from the COS-B γ -ray observations⁷. This latter number is an upper limit, since it includes a significant contribution from bremsstrahlung and inverse Compton γ -rays.

Various models have been designed to either increase the $\bar{p}$ lifetime τ , incorporate additional $\bar{p}$ sources, or conceal the π^0 γ -rays. Although the integral production rate of $\bar{p}$ can then be made to agree with observations, most models still fail to predict a substantial low-energy $\bar{p}$ flux because of the low energy cutoff that results from secondary interactions with stationary targets, in disagreement with the observation³. Models that provide deceleration of the $\bar{p}$ after production or require that the Solar System occupies a special position in the Galaxy have also been proposed^{8,9}.

We examine the possibility that the low-energy antiprotons observed in the cosmic radiation are produced through secondary p-p interactions in relativistic plasmas in the vicinity of neutron stars or black holes. At proton temperatures $\theta_p = kT/m_p c^2 \geq 0.2$ ($\theta_p = 1$ corresponds to $\sim 10^{13} \text{ K}$), a significant number of secondary $\bar{p}$ can be produced. We have calculated the rate coefficients and production spectra^{10,11} for secondary $\bar{p}$ using the invariant cross-section of Tan and Ng¹² and the results are shown in Figs 1 and 2. We have also calculated the rate and luminosity coefficients from secondary π^0 production data. From Fig. 1 we find that the efficiency for $\bar{p}$ production compared with π^0 production is greater than in the cosmic-ray case for $\theta_p \geq 0.5$, so the total π^0 -decay γ -ray luminosity will not

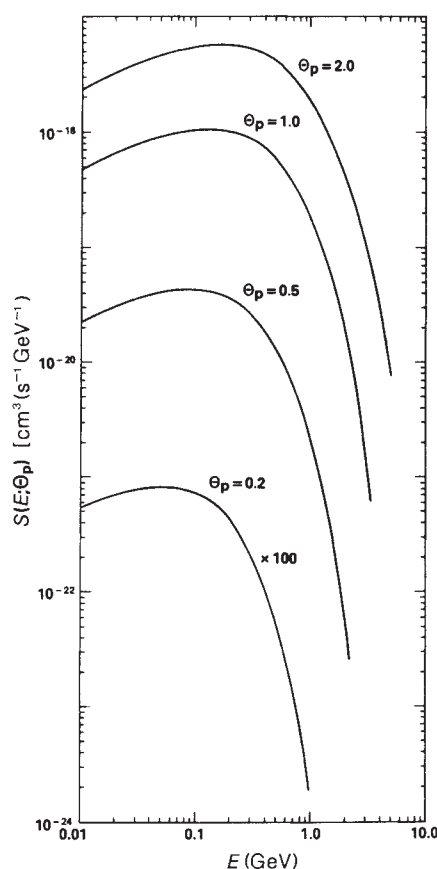


Fig. 2 Secondary $\bar{p}$ production spectra at various temperatures $\Theta_p = kT/m_p c^2$ are given as a function of kinetic energy E .

exceed limits implied by γ -ray observations of the Galaxy. In addition, several absorption processes likely to occur near compact objects, such as photon-photon or magnetic pair production, can further reduce the π^0 γ -ray luminosity.

The $\bar{p}$ spectra shown in Fig. 2 for a variety of temperatures extend to low energies without the appearance of the kinematic cutoff found in the cosmic-ray problem. The spectra peak at higher energies with increasing temperature, and exhibit an exponential decline above the peak temperature. But in all cases a very low-energy secondary $\bar{p}$ component is calculated, which could explain the low-energy $\bar{p}$ observation³.

Possible production sites of the $\bar{p}$ include the galactic bulge binary X-ray sources or the objects associated with the γ -ray point sources (for example, CygX-3). Models for disk accretion¹³ and spherical accretion incorporating shocks¹⁴ could yield proton temperatures as high as $\Theta_p \sim 1$. Although the establishment of a thermal distribution of protons may not be possible at these temperatures, this assumption affords the simplest calculations. Preliminary estimates employing nonthermal proton spectra in relativistic plasmas suggest low-energy $\bar{p}$ spectra similar in form to the results of Fig. 2.

Transport of the secondary $\bar{p}$ from the production site can occur because antineutrons $\bar{n}$, produced in equal numbers as the $\bar{p}$ (ref. 15), are not confined by the ambient plasma magnetic field. The $\bar{n}$ will escape to the interstellar medium and subsequently decay into $\bar{p}$ (the calculations of Figs 1 and 2 refer only to these particles). Because of the finite lifetime of the $\bar{n}$, an upper limit can be placed on the mass of black holes that serve as production sites for the $\bar{p}$. Requiring that the marginally relativistic $\bar{n}$ escape to ~ 10 Schwarzschild radii implies a maximum black hole mass of $< 10^7 M_\odot$. Past activity around such a massive black hole at the galactic centre could have produced the antiprotons without a large contemporaneous γ emission. A detailed comparison of $\bar{p}$ production in relativistic

plasma with observations must also include distortions of the emergent $\bar{n}$ due to the intense gravitational field of the compact object, reacceleration and energy losses of the source spectrum in the galactic environment, and modulation in interplanetary space.

We have presented a model that can produce low-energy antiprotons through secondary p-p interactions in relativistic plasmas. Such a model is in agreement with the observation of low-energy antiprotons in the cosmic radiation and the observed γ -ray luminosity of the galaxy. Moreover, it agrees with the present lack of observations of antinuclei in the cosmic radiation, whose formation by secondary production processes is entirely negligible.

C.D.D. is a NAS/NRC resident research associate.

Received 27 September; accepted 5 November 1985.

1. Golden, R. L. *et al. Phys. Rev. Lett.* **43**, 1196-1199 (1979).
2. Bogomolov, E. A. *et al. Proc. 16th Int. Cosmic Ray Conf.* **1**, 330-335 (1979).
3. Buffington, A. *et al. Astrophys. J.* **248**, 1179-1193 (1981).
4. Tan, L. C. & Ng, L. K. *J. Phys. G* **9**, 227-242 (1983).
5. Stecker, F. W. *et al. Astrophys. Space Sci.* **96**, 171-177 (1983).
6. Stephens, S. A. & Badhwar, G. D. *Astrophys. Space Sci.* **76**, 213-233 (1981).
7. Bloemen, J. B. G. M. *et al. Astrophys. J.* **279**, 136-143 (1984).
8. Stephens, S. A. & Mauger, B. G. in *Proc. 19th Recontre de Moriond Astrophysics Meet.*, 217-225 (1984).
9. Tan, L. C. & Ng, L. K. *Astrophys. J.* **269**, 751-764 (1983).
10. Weaver, T. A. *Phys. Rev. A* **13**, 1563-1569 (1976).
11. Dermer, C. D. *Astrophys. J.* **280**, 328-333 (1984).
12. Tan, L. C. & Ng, L. K. *Phys. Rev. D* **26**, 1179-1182 (1982).
13. Eilek, J. A. *Astrophys. J.* **236**, 664-673 (1980).
14. Meszaros, P. & Ostriker, J. P. *Astrophys. J.* **273**, L59-L63 (1983).
15. Gaisser, T. K. & Maurer, R. H. *Phys. Rev. Lett.* **30**, 1264-1267 (1973).

Specific role of transient $O^-(s)$ at Mg(0001) surfaces in activation of ammonia by dioxygen and nitrous oxide

C. T. Au & M. W. Roberts

Department of Chemistry, University College, Cardiff CF1 1XL, UK

It has been established¹⁻³ that oxygen chemisorbed at a metal surface can activate an otherwise unreactive molecule, as in the activation of S-H, N-H, O-H and C-H bonds. We have now extended these studies to more complex systems involving co-adsorbed molecules in the hope of throwing some light on the mechanisms of heterogeneous catalysts. We have previously established the mechanisms by which unreactive molecular water, adsorbed on a Zn(0001) surface, is activated when co-adsorbed with nitric oxide⁴, and have demonstrated N-H activation in ammonia co-adsorbed with nitric oxide at an atomically clean Mg(0001) surface at 295 K (Fig. 1a and ref. 5). Activation was attributed to surface oxygen generated in the dissociative chemisorption of nitric oxide. Here we establish the nature of the 'oxygen' species involved and reveal details of the molecular steps.

It is important to establish whether the 'oxygen' generated in the act of dissociative chemisorption of nitric oxide, that is, a short-lived species $O^-(s)$, has any specific chemical reactivity that distinguishes it from the thermodynamically more stable surface 'oxide' species, $O^{2-}(a)$, which was observed in the post-reaction XPS (X-ray photo-electron spectroscopy) study. Dynamic XPS over a wide temperature range is used to obtain information concerning the individual molecular steps.

We established first that a pre-oxidized Mg(0001) surface did not chemisorb ammonia at 295 K. Bulk MgO is also known to be unreactive to ammonia as judged by infrared spectroscopy^{6,7}. The active oxygen species is clearly not the thermodynamically stable $O^{2-}(a)$ species formed in the dissociative chemisorption of nitric oxide. However, dioxygen-ammonia mixtures also generate stable $NH_2(a)$ species⁸, which raises the issue of the possible steps that occur during the dissociative chemisorption of

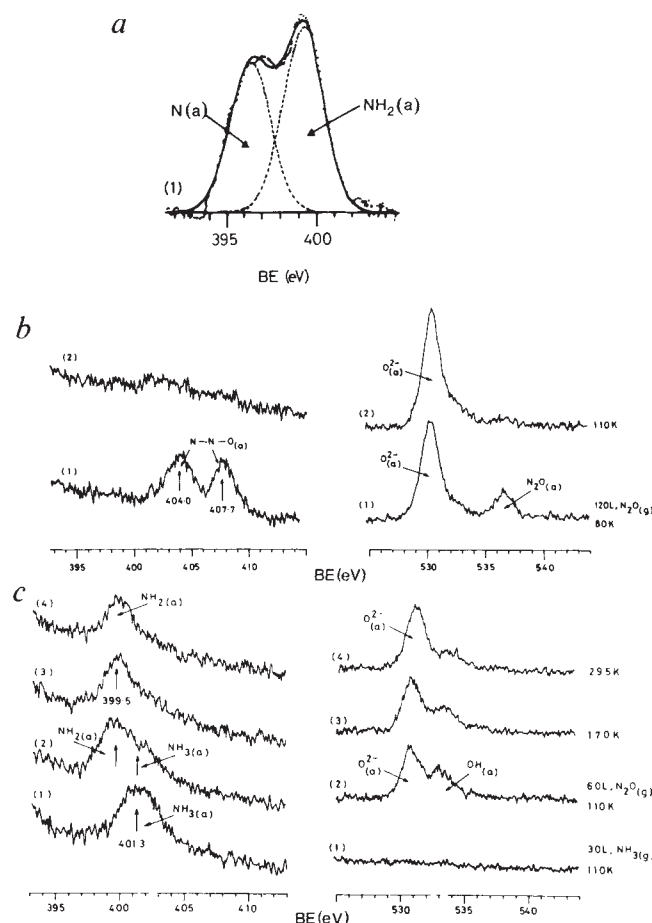
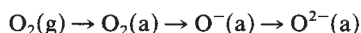


Fig. 1 *a*, Mg(0001) surface exposed to a mixture of NO(g) and NH₃(g) at 295 K (1L (langmuir) = 10⁻⁶ torr s). . . . , Raw N(1s) spectra after background removal and smoothing; ---, least-squares curve-fitted compounds assigned to NH₂(a) and N(a); BE, binding energy, *b*, N(1s) and O(1s) spectra for Mg(0001) surface exposed to N₂O(g) at 80 K (1) followed by warming the adlayer to 110 K (2). *c*, N(1s) (left) and O(1s) (right) spectra observed when Mg(0001) surface exposed to NH₃(g) at 110 K (10⁻⁶ torr 30 s) and then N₂O(g), 10⁻⁶ torr for 60 s, followed by warming to 170 and 295 K.

dioxygen at a metal surface and leading to oxidation. As kinetic (XPS) data have established the existence of a molecularly adsorbed oxygen precursor in the oxidation of aluminium at low temperatures (77 K) and both XPS and ultraviolet photoelectron spectroscopy have established⁴ the presence of O⁻(a) species during oxygen interaction with Zn(0001) surfaces, the following steps can be delineated for the dissociative chemisorption of oxygen at a metal surface



The step O⁻(a) → O²⁻(a) is initially kinetically fast at a Zn(0001) surface even at 77 K but with increasing O²⁻(a) concentration the process becomes activated so that at 77 K the O⁻(a) species represent a substantial (~60%) fraction of the O²⁻(a) concentration⁴. As the O⁻(a) species have been shown to abstract hydrogen from molecularly adsorbed water a mechanism of NH₃ activation could involve the intermediate O⁻(s). Note that there is no *a priori* reason why the above scheme for the dissociative chemisorption of dioxygen should not also include the O₂⁻(s) species and present at the surface as a transient. By studying the co-adsorption of ammonia with nitrous oxide starting at low temperature (110 K), this point can be resolved because N₂O is anticipated and shown to chemisorb (Fig. 1*b*) as indicated below. We use O⁻(s) to represent a surface transient and to

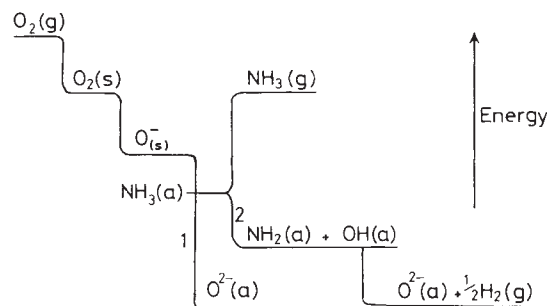
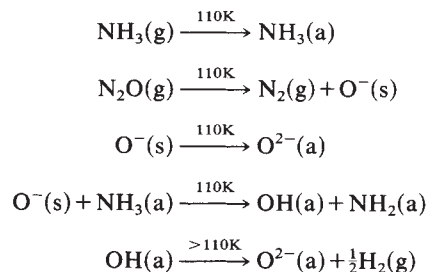


Fig. 2 Reaction scheme. Schematic representation of reaction pathways in the co-adsorption of 'oxygen' (N₂O, NO or O₂) and ammonia at 295 K; the ammonia diverts the 'oxygen' away from pathway 1 (the oxide route) and along pathway 2 (the amide route) in spite of its very weak interaction with Mg(0001); ammonia desorbs at 170 K. Also an oxide overlayer, O²⁻(a), is, like bulk MgO, unreactive to ammonia under these conditions.

distinguish it from the more stable O⁻(a) and O²⁻(a) species for which direct XPS evidences is available,



Figure 1*c* shows the N(1s) and O(1s) spectra observed when N₂O and NH₃ co-adsorbed at 110 K at a Mg(0001) surface was warmed to 295 K. Ammonia is physically adsorbed at 110 K and characterized by a single N(1s) peak at a binding energy of 401.3 eV. The calculated^{2,9} surface concentration is 0.4 × 10¹⁵ cm⁻². When this adlayer is exposed to N₂O(g), the N(1s) peak develops a second component at 399.5 eV. This is assigned⁹ to the NH₂(a) species. The O(1s) spectrum indicates surface oxide, O²⁻(a), and hydroxyl species with binding energies of 530.5 and 533.0 eV, respectively. On increasing the temperature first to 170 and then to 295 K, ammonia desorbs and intensity is lost from the 533.0-eV component to the 530.5-eV peak. The steps involved are as follows



There is, therefore, good experimental evidence that the oxygen species participating in the activation of ammonia is a transient, O⁻(s), formed during the oxidation of Mg(0001) by dioxygen, nitrous oxide and nitric oxide. We emphasize that the surface amide species NH₂(a) are strongly chemisorbed and could not have been anticipated on the basis of the lack of reactivity of NH₃ with both atomically clean Mg(0001) and with the 'bulk' MgO. Oxygen is, therefore, both a promoter and a poison for the activation; in its thermodynamically stable oxide state O²⁻(a), it is unreactive. We can envisage ammonia as an 'interceptor' molecule in that it influences the particular reaction pathway chosen by the co-adsorbed system. It is the efficiency of the NH₃—O⁻(s) step, that is, the ability of the ammonia to re-route the 'oxygen' along pathway 2 and away from pathway 1, that is crucial. This step can only be properly understood when knowledge of the surface lifetime of O⁻(s) becomes available. There may be alternative methods of generating the active oxygen species (such as lasers and discharges) so that unlikely reaction pathways, with (relatively) unfavourable thermodynamics can compete successfully with routes which at first sight,

that is, with no knowledge of the details of the molecular events involved, would be considered most probable.

There are also important analogies to be drawn between our results and those reported recently^{10,11} by Lunsford's group and commented on by Thomas¹² where activation of the C—H bond in methane by oxygen led to the formation of methyl radicals and surface hydroxyls. This was observed at high temperature (500 °C) with lithium-doped MgO where the active oxygen species was suggested to be O[•](s). The NH₂ species generated by hydrogen abstraction from ammonia at low temperature (Fig. 1) are also free radicals but, rather than desorb, are rapidly chemisorbed at the 'free' magnesium sites. Depending on the precise conditions at the surface, they may react with other species (if the number of Mg⁰ sites is limited), including each other, and desorb as hydrazine. By changing the 'oxygen'-to-ammonia ratio, evidence for this alternative possibility has been obtained⁸. Note that in Lunsford's work the addition of water vapour to the oxidant was beneficial¹⁰. We have established elsewhere^{9,13,15} that water vapour can enhance the presence of surface defects such as O[•](a) and OH(a) and which we have suggested⁹ accounts for some oxides such as NiO·OH exhibiting specific oxidation behaviour to organic molecules.

We thank SERC for support.

Received 19 August; accepted 4 November 1985.

1. Au, C. T. & Roberts, M. W. *Chem. phys. Lett.* **74**, 472–474 (1980).
2. Roberts, M. W. *Adv. Catalysis* **29**, 51–95 (1980).
3. Roberts, M. W. *Proc. Indian natn. Sci. Acad.* **51A**, 165–179 (1985).
4. Au, C. T. & Roberts, M. W. *Proc. R. Soc. A396*, 165–181 (1984).
5. Au, C. T., Carley, A. F. & Roberts, M. W. *Phil. Trans. R. Soc.* (in the press).
6. Tench, A. J. & Giles, D. *JCS Faraday Trans. I* **68**, 193–196; 197–201 (1972).
7. Coluccia, S., Garrone, E. & Borello, E. *JCS Faraday Trans. I* **79**, 607–613 (1983).
8. Au, C. T. & Roberts, M. W. (in preparation).
9. Au, C. T., Carley, A. F. & Roberts, M. W. *Int. Rev. phys. Chem.* (in the press).
10. Driscoll, D. J., Martir, W., Wang, Ji-Xiang & Lunsford, J. H. *J. Am. chem. Soc.* **107**, 58–63 (1985).
11. Ito, T. & Lunsford, J. H. *Nature* **314**, 721–722 (1985).
12. Thomas, J. M. *Nature* **314**, 669–670 (1985).
13. Carley, A. F., Grubb, S. & Roberts, M. W. *JCS chem. Commun.* 459–460 (1984).
14. Moroney, L., Roberts, M. W. & Smart, R. *JCS Faraday Trans. I* **79**, 1769–1778 (1983).
15. Carley, A. F., Rassias, S. & Roberts, M. W. *Surface Sci.* **135**, 35–51 (1983).

A new method for measuring palaeomagnetic intensities

Mark Boyd

Department of Earth Sciences, University of Cambridge,
Madingley Road, Cambridge CB3 0EZ, UK

Because measurements of palaeomagnetic intensities (henceforth palaeointensities) are usually time-consuming and the success rate only poor, say, 30–50%, many palaeointensity values are based on as few as three experimental results. In an attempt to improve on this, a technique has been developed using automated equipment which is faster, and which can detect alteration of a sample during measurement, allowing the experiment to be terminated and another sample to be tried. The technique, described here, has been used on a collection of ancient bricks, and the results show the palaeofield at Cambridge (England) to have decreased from about 49 μ T in AD 1511 to 42 μ T in AD 1758, in good agreement with results using Shaw's method.

A SQUID (superconducting quantum interference device) magnetometer is used to measure the magnetic moment of a 4 mm diameter sample held at an elevated temperature within a compact non-magnetic oven that can be heated and cooled rapidly because of its low thermal capacity. Because the magnetometer is totally automated and the sample does not have to be repeatedly cooled to room temperature, as in the Thellier¹ and similar methods, palaeointensities can be obtained in

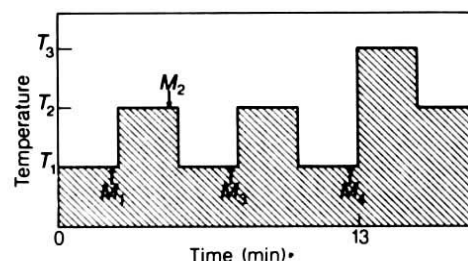


Fig. 1 Sequence of steps for one palaeointensity measurement typically a sample spends 130 s at equilibrium at each temperature.

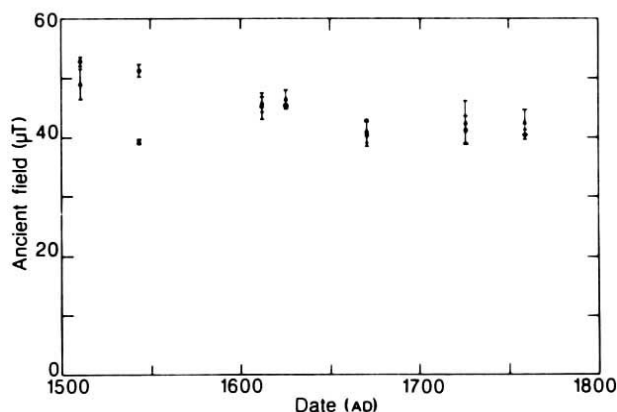


Fig. 2 Intensity of the geomagnetic field at Cambridge since AD 1500. ● Results from Shaw method; ▲, results from the present thermal method. Each point is the average result from three samples and the error bars one standard deviation. Only the AD 1543 bricks show significant disagreement.

75 min; Shaw's method², by contrast, takes at least 140 min and the Thellier method can take anything from 3 h to several days.

The measurement sequence which allows the calculation of a palaeointensity (Fig. 1) begins by heating the sample to a base temperature T_1 and taking the first measurement of its magnetic moment M_1 . Raising the temperature to T_2 , a second measurement, M_2 , is made before the sample is cooled back to the base temperature for the M_3 measurement, thermally demagnetizing the sample. At this point in the sequence, a magnetic field of known strength and direction is applied to the sample and maintained while the sample is again cycled to temperature T_2 and cooled to T_1 , giving the sample a partial thermal remanent magnetization (PTRM). After removing the magnetic field the final measurement of the sequence M_4 is made.

This procedure effectively selects the magnetic grains which have blocking temperatures between T_1 and T_2 . The natural remanent magnetization (NRM) preserved in this group of grains is the difference between M_1 and M_3 whereas the difference between M_4 and M_3 is the magnitude of the PTRM that these grains have acquired in the laboratory field. The ratio of these values combined with the laboratory field strength F_L gives the ancient field intensity, F_A

$$F_A = F_L \times \frac{M_1 - M_3}{M_4 - M_3}$$

To examine the grains with blocking temperatures between T_2 and T_3 the measurement cycle is repeated raising the base temperature to T_2 . Because the M_2 measurement from the previous cycle is identical to M_1 for the present one, the initial measurement can be omitted from all subsequent cycles. Thus, six or seven cycles can be made before the Curie temperature of the sample is reached.

Although the sample quickly reaches thermal equilibrium after each heating and cooling phase, its magnetic moment is

Table 1 Two complete sets of experimental results

Temperature	M_1	M_3	M_4	M'_2	Ratio $\frac{M_1 - M_3}{M_4 - M_3}$	Alteration estimate $\frac{(M'_2 - M_2)}{M_1 - M_3} \times \frac{M_3}{M_2}$
No alteration						
110	11,546	10,918	11,669		0.84	-1.8
165	10,219	8,951	10,289	10,208	0.95	0.8
230	8,169	7,235	8,349	8,178	0.84	-1.2
310	6,348	5,202	6,434	6,337	0.93	1.3
375	4,356	3,252	4,554	4,369	0.85	3.0
440	2,502			2,528		
Alteration >7%						
110	3,922	3,658	4,013		0.74	6.6
165	3,459	3,308	3,671	3,476	0.42	20.4
230	3,137	3,054	3,337	3,166	0.29	0.4
310	2,829	2,672	3,039	2,830	0.43	30.9
375	2,413	2,195	2,612	2,457	0.52	-20.4
440	1,930			1,890		

The top set is from a sample which suffers no alteration. The ancient field intensities predicted at each temperature step are consistent. The lower set of results have values of >7% alteration effect, consequently this sample was rejected.

far from stable. Viscosity effects, the relaxation of grains with blocking temperatures higher than the sample temperature, mean that its magnetic moment is constantly changing. However, these viscous changes of moment can be allowed for and good results obtained provided. (1) each phase of the experiment, after the sample has reached a new thermal equilibrium, lasts the same length of time, typically 130 s; and (2) all measurements of the moment are taken at exactly the same length of time after thermal equilibrium is achieved. Usually the moment is averaged over the period 100–130 s.

Care must be taken that chemical or physical alteration of the magnetic carriers during measurement does not produce apparently acceptable, yet erroneous, results. This method provides several independent tests for such alteration. (1) The ratios obtained from each temperature step should be the same—this is equivalent to the linearity test used by the Thellier method. (2) Measurements of the samples' susceptibility are repeated at each base temperature which may show up any alteration. (3) The difference between M_2 and M_3 is caused by the increase in spontaneous magnetization of the oriented grains with blocking temperature above the temperature at which M_2 was taken. If an additional cooling step is included at the end of each cycle the saturation magnetization change can again be measured, checking for alteration that may occur during the heating phases between the two measurements. (4) A final heating step can be added to each cycle to repeat the M_2 measurement and monitor any alteration that may occur during the cycle. This will not show any alteration that has occurred during heating immediately before M_2 . However, it is unlikely that this alteration will have saturated; it will continue while the sample is held at this temperature. The repeat measurement M'_2 should record the resultant change of moment which is expressed as a percentage of NRM lost in this single temperature cycle. Table 1 shows two sets of experimental data, the first sample did not suffer alteration but the second had to be discarded. An alteration effect of >7% is considered unacceptable.

Tests (3) and (4) show up alteration occurring only in the magnetic grains whose moment is actually contributing to the estimate of the ancient magnetic field. Alteration of the other magnetic grains is tolerable as they do not influence the ancient field estimate. Such alteration may be detected by conventional tests and cast unnecessary doubts on the results.

Four test samples were laboratory fired in a known field, using the above procedure in an attempt to recover the TRM field strength. The field strengths predicted ranged between +3.7 and -3.4% of the actual TRM field strength (Table 2).

Table 2 Results from four samples fired in a known magnetic field in the laboratory

Sample	TRM field (μ T)	Field during experiment (μ T)	Calculated cooling field (μ T)	Age error (%)
1	67.56	61.93	67.32 $\pm$ 0.83	-0.4
2	33.78	39.41	32.63 $\pm$ 1.19	-3.4
2	33.78	33.78	33.68 $\pm$ 0.74	-0.3
3	36.60	33.78	35.60 $\pm$ 0.64	-2.7
4	59.38	53.73	61.60 $\pm$ 1.58	+3.7
4	50.86	48.03	52.35 $\pm$ 1.17	+2.9

The method has recovered these fields to within an average of 2%.

To test this method, it was used on a collection of dated bricks taken from Colleges in the Cambridge area. These samples were also measured in the Cardiff laboratory⁴ using the modified Shaw method⁵. Figure 2 shows that, except for one brick both sets of results were in good agreement. The average error between the two sets is 3.4%, four results being higher with the thermal method and two lower. These indicate that the geomagnetic field intensity at Cambridge has decreased smoothly from ~49 μ T in AD 1511 to 42 μ T in AD 1758. The original brick collection contained samples dated up to 1979, but all the bricks younger than 1758 exhibited unacceptable alteration and produced no results.

The major advantage of the thermal palaeointensity method described here is that its speed allows more samples to be measured so that the high failure rate, common to all thermal methods, becomes more tolerable. More results allow more accurate determinations of the ancient field. In addition, this experimental procedure automatically gives information on the spontaneous magnetization curve of the sample and a measurement of its d.c. susceptibility at each base temperature. This information in surplus to that required to produce an estimate of the ancient magnetic field but may give an indication of the magnetic carriers present in the sample.

This work was funded by N.E.R.C. grant GR3/4670 and the Northern Ireland Education Board.

Received 5 August; accepted 30 October 1985.

1. Thellier, E. & Thellier, O. *Annls Geophys.* **15**, 285–376 (1959).
2. Shaw, J. *J. Geophys. J.R. astr. Soc.* **39**, 133–144 (1974).
3. Walton, D. *Nature* **310**, 740–743 (1984).
4. Shaw, J., Share, J. A. & Rogers, J. *Geophys. J.R. astr. Soc.* **78**, 209–217 (1984).
5. Rolph, T. C. & Shaw, J. *Geophys. J.R. astr. Soc.* **80**, 773–781 (1985).

Seismic reflection imaging of the subducting Juan de Fuca plate

A. G. Green*, R. M. Clowes†, C. J. Yorath‡, C. Spencer*, E. R. Kanasewicz§, M. T. Brandon‡ & A. Sutherland Brown‡

* Division of Seismology and Geomagnetism, Department of Energy, Mines and Resources, 1 Observatory Crescent, Ottawa, Ontario, Canada K1A 0Y3

† Department of Geophysics and Astronomy, University of British Columbia, Vancouver, British Columbia, Canada V6T 1W5

‡ Pacific Geoscience Centre, Department of Energy, Mines and Resources, Sidney, British Columbia, Canada V8L 4B2

§ Department of Physics, University of Alberta, Edmonton, Alberta, Canada T6G 2E1

McKenzie and Parker's early suggestion¹ that the Juan de Fuca plate is underthrusting North America has since been confirmed by numerous studies. Evidence for plate convergence along this junction includes: a sediment-filled trench²⁻⁷, intense compressional deformation of the continental slope and shelf sediments²⁻⁷, seafloor magnetic anomalies that extend beneath the continental slope^{2,4}, inland andesitic volcanoes⁸, paired low/high heat flow and gravity anomalies^{9,10}, active onshore deformation^{11,12}, onshore crustal earthquakes with compressional axes parallel to the direction of plate convergence¹³, sinistral strike-slip earthquakes along the Nootka Fault (Fig. 1)¹⁴, and a well-defined Benioff-Wadati zone^{6,15,16}. Here we report the results of a seismic reflection survey that has delineated two slabs of oceanic lithosphere underlying Vancouver Island, one that is currently being subducted and one

that is underplated. These findings lead us to speculate that successive underplating of oceanic lithosphere may be an important process in the evolution and growth of continents.

A total of 205 km of deep seismic reflection data have been collected along the four profiles shown in the simplified geological map of southeastern Vancouver Island (Fig. 1). The VISP1 line crosses the width of the island and is almost coincident with one of the gravity profiles studied by Riddihough¹⁰ and with the combined onshore/offshore seismic refraction line of Spence *et al.*¹⁷. Some three-dimensional control on the interpretation of VISP1 is available from the VISP3 line located 15–20 km to the east and from a short test line described by Clowes *et al.*¹⁸. The two southeasterly lines, VISP2 and VISP4, were designed to determine the attitudes and significance of the San Juan, Survey Mountain and Leech River faults.

Line drawings of the reflections are shown in Fig. 2 and a typical example of the data, with an accompanying simplified interpretation, is shown in Fig. 3. Two very prominent reflection zones, C and E in Figs 2 and 3, are observed on all seismic sections. They underlie most of southeastern Vancouver Island including the youngest allochthonous terrains along the southeastern tip of the island. Both reflection zones dip to the north-east, parallel to the direction of convergence between the Juan de Fuca and North American plates (Fig. 1).

Reflection zone E probably originates from the boundary region between the descending Juan de Fuca plate and the overriding North American plate. It lies at a depth of only ~23 km beneath the southern end of VISP1, based on migrated travel-times and velocities from Spence *et al.*'s¹⁷ seismic refraction model, and is deepest at ~34 km beneath the north-central region of the island. Its overall dip in the vicinity of VISP1 is to the north-east at an angle of 9–13°. From Fig. 2, the average

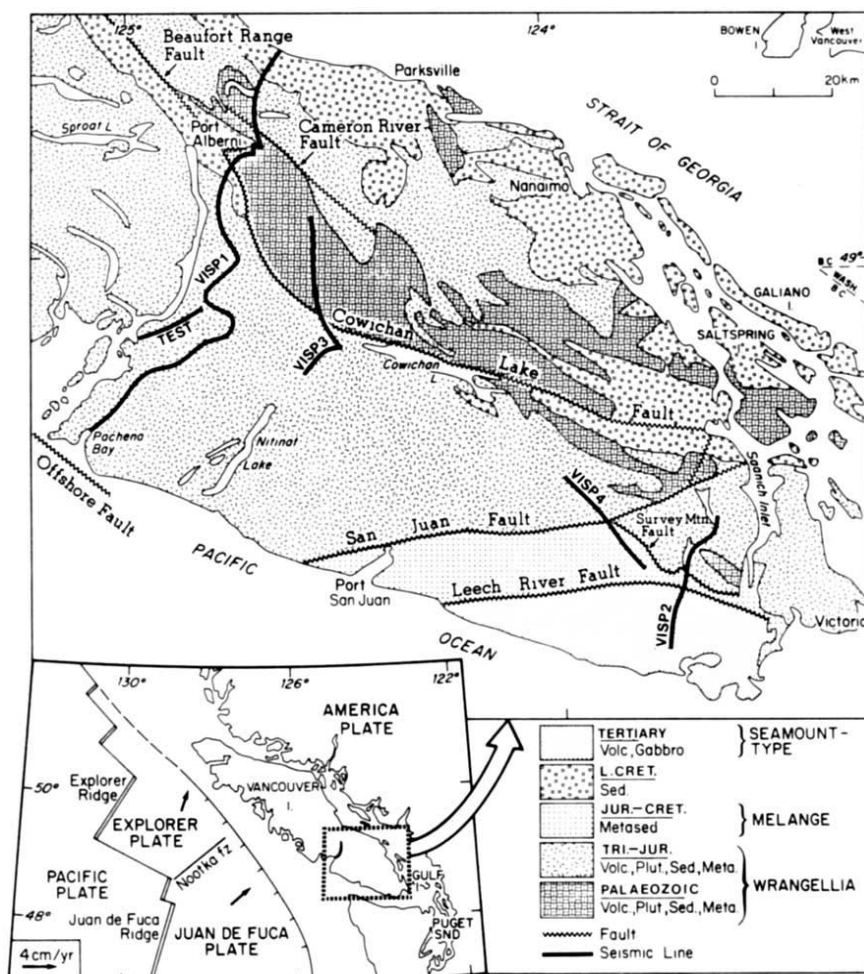


Fig. 1 Simplified geological map of southeastern Vancouver Island showing the locations of the four seismic reflection profiles, VISP1 to VISP4, and the short seismic reflection profile, TEST, of Clowes *et al.*¹⁸. Geology is mostly from Muller¹⁴, with minor modifications based on recent mapping programmes. The region to the north of the San Juan Fault is part of the exotic Wrangellia terrain of Palaeozoic and Mesozoic age. Between the San Juan and Leech River faults is an allochthonous mélange of Mesozoic metasediments and to the south of the Leech River Fault lies a volcanic terrain that may have been part of a Cenozoic seamount chain. The inset shows the location of the geological map with respect to the major plate boundaries.

depth to zone E is shown to be somewhat less than the depth to the active subduction zone of the seismic refraction model. However, Spence *et al.*¹⁷ have demonstrated that the depth and structure of the subducting plate beneath Vancouver Island are only weakly constrained by the onshore/offshore seismic data. The location and attitude of zone E match well Riddihough's¹⁰ estimates for the top of the subduction zone based on a variety of seismic and gravity data, and the northeasterly dip of 9–13° is in excellent accord with the ~9° value obtained from an onshore/offshore seismic refraction study of the Washington continental margin¹⁹ and the 11–15° dip of the Benioff–Wadati zone below southern Vancouver Island, Puget Sound and western Washington^{6,15,16}. The shallow dip estimates are also consistent with the pattern of recent onshore deformation in this region¹¹ and with teleseismic waveforms recorded at seismographic stations on Vancouver Island and in western Oregon²⁰.

Towards the northern end of VISP1 there is some evidence for an increase in the dip of reflection zone E (Fig. 2). It is near to this location that the subducting plate in Riddihough's density models¹⁰ begins to plunge more steeply into the mantle. Such a feature is common in other subduction zones²¹ and is generally consistent with the steeply dipping slab of high-velocity material

required to explain the pattern of travel-time anomalies to the north-east of Puget Sound²². The apparent absence of the deepest reflections beneath the northernmost region of VISP1 has at least two possible explanations. This part of the profile traverses the nose of a NW–SE-trending anticlinal ridge that exposes the oldest rocks on the island before crossing an escarpment onto a late Cretaceous sedimentary basin (Fig. 1), so the absence of deep reflections here may be related to geological and topographic complexities. Alternatively, if the dip of the subducting plate increases to a value in excess of ~45°, as shown in some of Riddihough's density models¹⁰ and as required by McKenzie and Julian's travel-time analysis²², the relatively conventional data processing that we have applied would have effectively filtered out any reflections from this zone.

Figure 4 shows a contour map of two-way travel-times to the top of reflection zone E. Particularly noteworthy is the structure that is required to explain the variations in depth and attitude of the reflection zone between the western pair of seismic profiles (1 and 3) and the eastern pair (2 and 4). Of course, the dashed parts of the contour diagram are speculative; a major fault anywhere between the two pairs of profiles would allow an equally valid interpretation of existing data. Warping and frac-

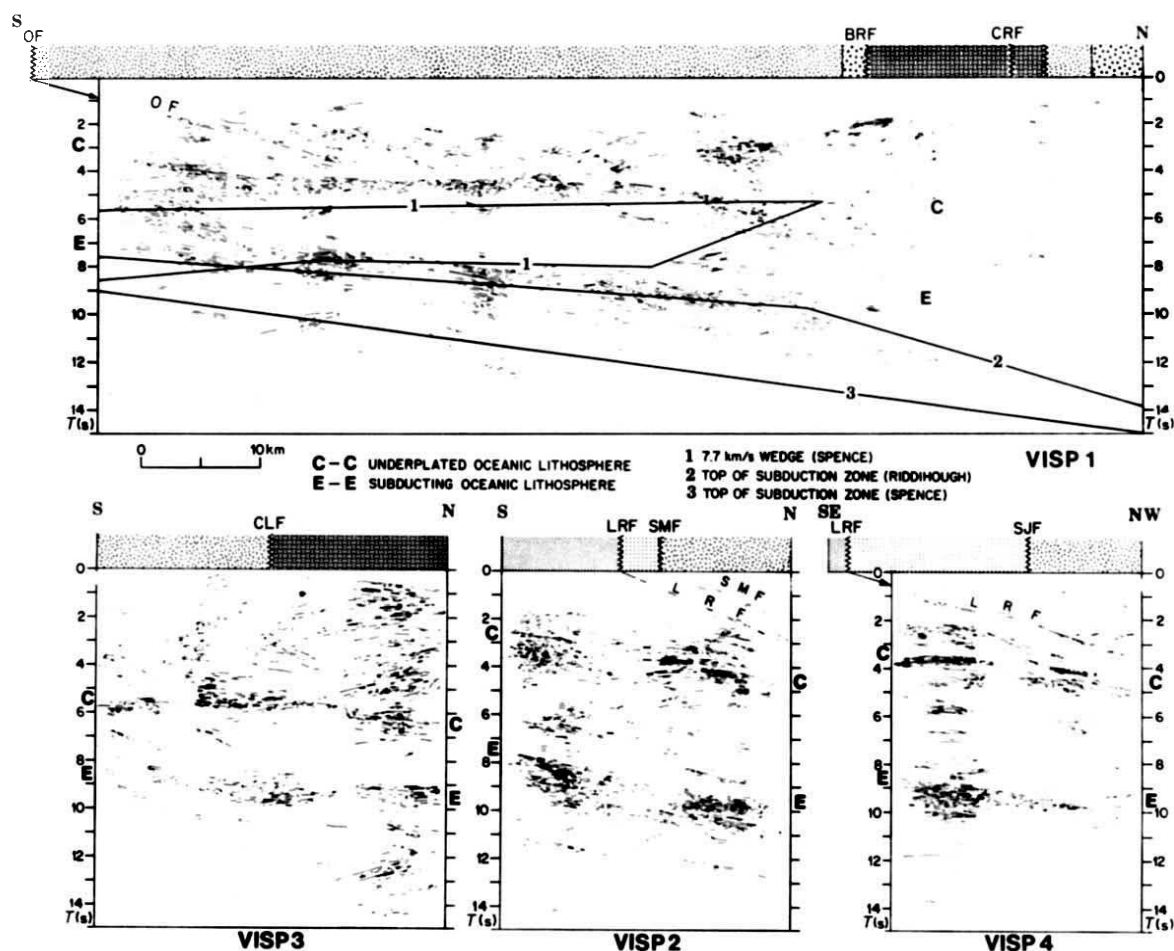


Fig. 2 Line drawings showing the unmigrated reflections recorded on the four seismic reflection profiles. Vertical scale is two-way travel-time in seconds (T). All lines were recorded to 16 s with 30-fold coverage. Acquisition methods were similar to those used in the US by the COCORP group⁴⁵. Final processing included single-fold trace editing, crooked line geometry and elevation corrections, automatic gain control, normal move-out corrections, trim statics using a correlation window of $T = 1$ –12 s, common reflection point stacking, bandpass filtering of 8–40 Hz and trace-to-trace amplitude equalization using a $T = 4$ –8 s window. Patterns at the top of each section show the surface geology crossed; an explanation of the various patterns is given in Fig. 1. On the VISP1 section the -1- border delineates the high-velocity (7.7 km s^{-1}), high-density (3.3 g ml^{-1}) block proposed by Spence *et al.*¹⁷, and the -2- and -3- boundaries show, respectively, the Riddihough¹⁰ and Spence *et al.*¹⁷ estimates for the top of the actively subducting Juan de Fuca plate. C–C and E–E are prominent reflection zones that delineate the upper boundary of the high-velocity, high-density block (the underplated oceanic lithosphere) and the top of the actively subducting Juan de Fuca plate respectively. OF, Offshore (or Tofino⁴⁶) Fault; BRF, Beaufort Range Fault; CRF, Cameron River Fault; CLF, Cowichan Lake Fault; LRF, Leech River Fault; SMF, Survey Mountain Fault; SJF, San Juan Fault (locations of faults are shown in Fig. 1). A typical example of the seismic reflection data is shown in Fig. 3.

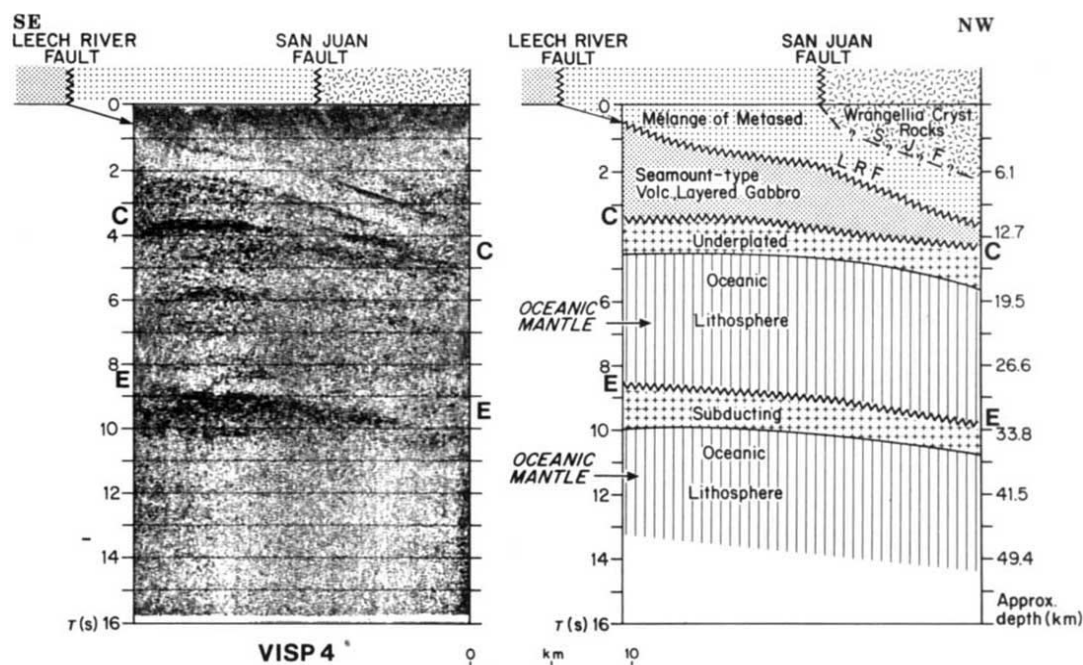


Fig. 3 A typical unmigrated seismic reflection section and simplified interpretation obtained from data recorded on Vancouver Island. Figure 1 shows the location of the profile and Fig. 2 (bottom right) shows a line diagram of the individual reflections recorded on the section. Patterns at the top of both the section and the interpretation diagram show the surface geology; an explanation of the various patterns is given in Fig. 1. The regions of + signs correspond to the reflection zones C and E. Approximate depths shown are based on the seismic refraction model of Spence *et al.*¹⁷. The seismic reflection data were recorded by Veritas Geophysical Ltd and processed by Veritas Seismic Ltd.

turing of the subducting plate in this general region have been predicted by Rogers¹⁵ and by Keen and Hyndman⁹ on the basis of seismicity patterns, fault plane solutions and the geometry of the Juan de Fuca–North American plate interactions. Some relatively deep strike-slip earthquakes in western Washington suggest that intra-plate faulting or segmentation may be a common feature of the downgoing Juan de Fuca plate¹⁶.

The nature of the E reflection zone requires some discussion. On the migrated sections it is 3–5 km thick (1–1.5 s two-way travel-time) and has a strongly laminated character that could result from layered sediments, intercalated volcanics and sediments, layered igneous rocks or tectonic structures induced by underthrusting. Undeformed sediments are known to have been subducted to considerable depths beneath the forearc basins of several island arcs, including those of Japan^{23,24}, the Aleutians^{25,26} and the Caribbean^{27,28}. However, in the general region of Vancouver Island there are numerous single- and multi-fold seismic reflection profiles which show that the majority of sediments have been scraped off the descending oceanic plate to form the accretionary wedge that now constitutes the continental margin^{2–7}. Reflection zone E is therefore unlikely to have originated from a simple stack of sedimentary layers riding on top of the descending oceanic plate. Nevertheless, it is feasible that the lowermost layers of lithified sediments have been taken down into the subduction zone and that these sediments and possibly the upper crystalline layers have been transferred or underplated to the base of the overriding plate. Successive underplating of this type would result in the formation of a thick layered sequence between the descending Juan de Fuca slab and the overriding North American plate. This particular process of accretion, termed subcretion by Karig and Kay²⁹, has been invoked by Von Heune²⁵ to explain some multi-fold seismic reflection data collected across the Aleutian trench and forearc basin.

Episodic and rapid sedimentation from the North American continent^{2–7}, combined with episodic volcanism involving extensive lava flows from the nearby oceanic spreading centres³⁰, have probably resulted in the widespread interlayering of oceanic basalts and sediments on the Juan de Fuca plate. Comprehensive studies across the northern parts of the Explorer plate (Fig. 1) and adjacent ridge segments have revealed ubiquitous interfingering of basaltic lavas and turbidites^{3,30–33}, and similar features are shown on the generalized geological

cross-section of the central Oregon continental margin³⁴.

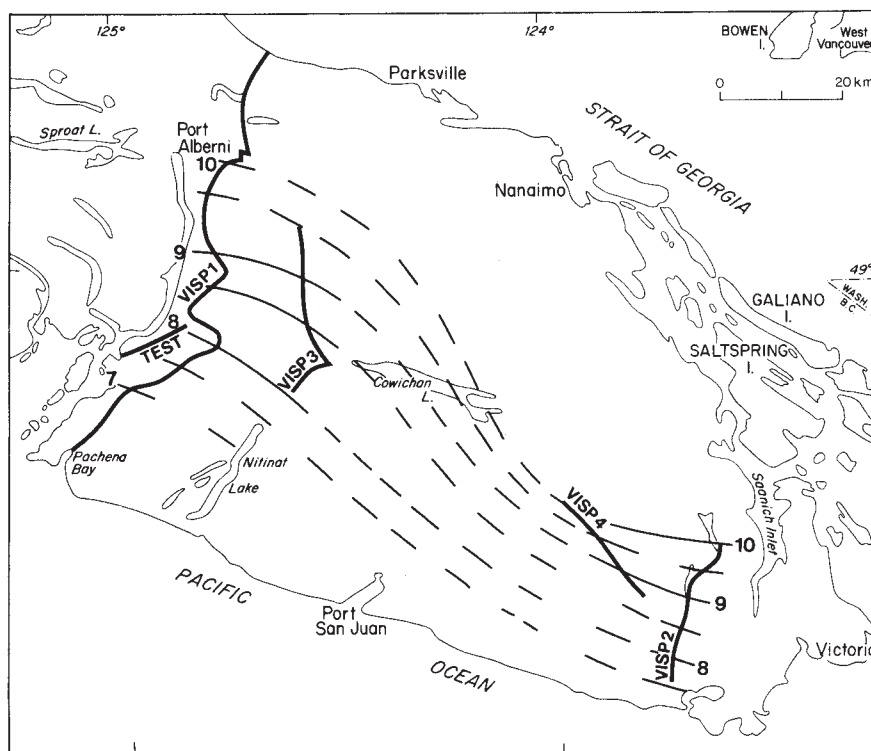
Based on the results of multi-fold seismic reflection surveys, high-quality seismic refraction surveys and deep-sea drilling, Talwani *et al.*²⁷ have proposed that interlayered extrusive and intrusive (mostly sills) rocks are the source of some laterally extensive reflections from within the acoustic basement of the small Caribbean plate. As similar lava flows and sills are expected to occur in the upper regions of the Juan de Fuca plate, some of the reflections in the E zone, particularly those near to the base of the zone, could have originated from structures within the descending oceanic crust.

Finally, at some subduction zones, notably those off the coasts of Java and Ecuador²⁴, there is a hint that layered reflections along the boundary of the descending plate only become pronounced landward of the trench. Layering of this nature could either be produced by the selective slicing away of sedimentary rocks from the underside of the accretionary wedge or it could result from brittle and/or ductile deformation induced by the relative motion between the two plates.

We interpret the region between reflection zones C and E as a tectonically underplated slab of older oceanic lithosphere, which was accreted to the base of Vancouver Island after the westward jump in the locus of subduction in Eocene times^{35,36}. As shown in Fig. 2, our interpretation is supported by the gravity and seismic refraction models which require high-density (3.3 g ml^{-1}) and high-velocity (7.7 km s^{-1}) mantle-type material and velocity discontinuities at the appropriate shallow depths. The upper boundary of the high-velocity material at $\sim 18 \text{ km}$ depth is well-delineated from interpretations of the combined onshore/offshore seismic refraction line¹⁷ and an intersecting refraction line that was recorded along the length of the island³⁷, so its coincidence with the base of reflection zone C is significant. On the seismic reflection lines the thickness of the interval between zones C and E varies from 9 to 18 km, which is comparable to the 17–20 km range that Spence *et al.*¹⁷ derived for the thickness of the high-velocity mantle material in the descending Juan de Fuca plate.

Reflection zone C is strikingly similar to reflection zone E. It has a similar laminated character, an average thickness of 3–5 km and an overall dip to the north-east of $5\text{--}8^\circ$. It seems to have acted as a decollement zone to a number of major northeasterly-dipping structures, most of which can be projected to the surface where they coincide with mapped faults or plutonic contacts.

Fig. 4 Contours are two-way travel-times (unmigrated) (in s) to the top of reflection zone E. The curvature of the dashed contours between the western pair of seismic profiles (VISP1 and VISP3) and the eastern pair (VISP2 and VISP4) is speculative; a major fault anywhere between the two pairs of profiles would be an equally valid interpretation of the travel-time information.



Relative motion during the waning stages of this earlier phase of plate convergence would have occurred along these moderately dipping structures and along the decollement zone. A necessary consequence of our tectonic model is the removal of the lower lithosphere of the overriding plate during Eocene and earlier times.

It seems to us that under favourable but frequently occurring conditions, tectonic underplating of oceanic lithosphere at the base of convergent margins is a rapid means of adding large volumes of material to terrains that eventually become continental. This process is capable of creating root zones beneath the mountain ranges of active continental margins and of substantially increasing the crustal thicknesses of island arc-trench systems. In addition to our Vancouver Island example, thin slabs of underplated oceanic lithosphere have been proposed to underlie parts of Java³⁸, the Aleutians²⁵ and southern Alaska³⁹. The underplated oceanic lithosphere would generally be of higher density, higher velocity and more basic than the overlying material. Across such terrains now interior to the continents, seismic refraction methods would resolve a broad-scale subhorizontal layering of the crust (including zones of high and low velocity), whereas seismic reflection techniques would see the pronounced reflection zones. Although many regions do not reveal such a simple image of the continental crust, a two- to four-layered crustal section with laminated reflections in the mid- to lower regions is a common feature on most continents⁴⁰. Of course, we do not claim that underplating is the only or the dominant means of crustal thickening, but it is an important process that requires further investigation. In particular, the new mechanism could help to explain the anomalously high continental growth rates obtained for Precambrian shields^{41,42}. Tectonic processes in Archaean times were probably dominated by the rapid movements of small, short-lived oceanic plates^{41,43} and, like the present situation along the west coast of North America, subduction of the young and buoyant oceanic lithosphere would have occurred at relatively shallow angles.

We thank Drs J. Adams, M. Berry and R. Kurtz for critical reviews of the first draft of this manuscript and Dr E. Irving for pointing out the significance of Reymer and Schubert's work. This work was conducted as part of the LITHOPROBE program

to study the three-dimensional structure of key geological targets on the Canadian landmass and adjacent offshore margins. We are grateful to Veritas Geophysical Ltd and Veritas Seismic Ltd for collecting and processing our data. This is LITHOPROBE paper 7; Earth Physics Branch contribution 1202.

Received 5 July; accepted 21 October 1985

- McKenzie, D. P. & Parker, R. L. *Nature* **216**, 1276-1280 (1967).
- Silver, E. A. *Mar. Geol.* **13**, 239-249 (1972).
- Tiffin, D. L., Cameron, B. E. B. & Murray, J. W. *Can. J. Earth Sci.* **9**, 280-296 (1972).
- Barr, S. M. *Can. J. Earth Sci.* **11**, 1187-1199 (1974).
- Ellis, R. M. *et al.* *Can. J. Earth Sci.* **20**, 719-741 (1983).
- Kulm, L. D. *et al.* *Atlas 1 Ocean Margin Drilling Program* (Marine Science International Regional Atlas Ser., 1984).
- Clowes, R. M. *et al.* *J. Can. Soc. exp. Geophys.* **20**, 23-39 (1984).
- Dickinson, W. R. *Rev. Geophys.* **8**, 813-860 (1970).
- Keen, C. E. & Hyndman, R. D. *Can. J. Earth Sci.* **16**, 712-747 (1979).
- Riddihough, R. P. *Can. J. Earth Sci.* **16**, 350-361 (1979).
- Savage, J. C., Lisowski, M. & Prescott, W. H. *J. geophys. Res.* **86**, 4929-4940 (1981).
- Adams, J. *Tectonics* **3**, 449-472 (1984).
- Weaver, C. S. & Smith, S. W. *J. geophys. Res.* **88**, 10371-10383 (1983).
- Rogers, G. C. *Can. J. Earth Sci.* **16**, 523-531 (1979).
- Rogers, G. C. *U.S. geol. Surv. Open File Rep.* **83-19**, 19-39 (1983).
- Taber, J. J. & Smith, S. W. *Bull. seism. Soc. Am.* **75**, 237-249 (1985).
- Spence, G. D., Clowes, R. M. & Ellis, R. M. *J. geophys. Res.* **90**, 6754-6772 (1985).
- Clowes, R. M. *et al.* *Nature* **303**, 668-670 (1983).
- Taber, J. J. thesis, Univ. Washington, Seattle (1983).
- Langston, C. A. *J. geophys. Res.* **86**, 3857-3866 (1981).
- Pennington, W. D. *Science* **220**, 1045-1047 (1983).
- McKenzie, D. P. & Julian, B. *Bull. geol. Soc. Am.* **82**, 3519-3524 (1971).
- Aoki, Y., Tamano, T. & Kato, S. *Mem. Am. Ass. Petrol. Geologists* **34**, 309-322 (1982).
- Lehner, P. *et al.* *Am. Ass. Petrol. geol. Stud. Geol. Ser.* **15** Vol. 3, 3.4.2.1-3.4.2.128 (1983).
- Von Heune, R. *Mem. Am. Ass. Petrol. Geologists* **29**, 261-272 (1979).
- Von Heune, R. *et al.* *Am. Ass. Petrol. Geol. Stud. Geol. Ser.* **15** Vol. 3, 3.4.2.135-3.4.2.139 (1983).
- Talwani, M. *et al.* *Am. Geophys. Un., M. Ewing Ser.* **1**, 83-98 (1977).
- Westbrook, G. K. *et al.* *Nature* **300**, 625-628 (1982).
- Karig, D. E. & Kay, R. W. *Phil. Trans. R. Soc. A303*, 233-251 (1981).
- Davis, E. E. *Bull. geol. Soc. Am.* **93**, 1023-1029 (1982).
- Clowes, R. N. & Knize, S. *Can. J. Earth Sci.* **16**, 1265-1280 (1979).
- Riddihough, R. P., Currie, R. G. & Hyndman, R. D. *Can. J. Earth Sci.* **17**, 577-593 (1980).
- Davis, E. E. & Riddihough, R. P. *Can. J. Earth Sci.* **19**, 767-788 (1982).
- Snively, P. D., Wagner, H. C. & Lander, D. L. *Bull. geol. Soc. Am.* **91**, 143-146 (1980).
- Dickinson, W. R. *Can. J. Earth Sci.* **13**, 1268-1287 (1976).
- Muller, J. F. *Can. J. Earth Sci.* **9**, 2062-2085 (1977).
- McMechan, G. A. & Spence, G. D. *Can. J. Earth Sci.* **20**, 742-752 (1983).
- Hamilton, W. *U.S. Geol. Surv. Prof. Pap.* 1078 (1979).
- Ambos, E. L. *et al.* *EOS* **66**, 301 (1985).
- Oliver, J. E., Cook, F. & Brown, L. *J. geophys. Res.* **88**, 3329-3347 (1983).
- Dewey, J. F. & Windley, B. F. *Phil. Trans. R. Soc. A303*, 189-206 (1981).
- Reymer, A. & Schubert, G. *Tectonics* **3**, 63-77 (1984).
- Abbot, D. H. & Hoffman, S. E. *Tectonics* **3**, 429-448 (1984).
- Muller, J. E. *Geol. Surv. Can. Open File Map* 463 (1977).
- Schilt, S. *et al.* *Rev. Geophys. Space Phys.* **17**, 354-368 (1979).
- Brandon, M. T. *Geol. Soc. Am. Cord. Sect. Mtg.* Vancouver, BC, 7.1-7.2.8 (1985).

Thermal expansion of silicate perovskite and stratification of the Earth's mantle

Elise Knittle*, Raymond Jeanloz* & Gordon L. Smith†

* Department of Geology and Geophysics, University of California, Berkeley, California 94720, USA

† Lawrence Livermore National Laboratory, Livermore, California 94550, USA

A pressure of 24 GPa and temperatures of about 2,000–3,000 K correspond to the observed 670-km seismic discontinuity which separates the upper and lower mantle. In such conditions the major minerals of the Earth's upper mantle, olivine ((Mg, Fe)₂SiO₄), pyroxene ((Mg, Fe)SiO₃) and garnet ((Mg, Fe, Ca)₃Al₂Si₃O₁₂) transform to a distorted (orthorhombic) perovskite-structured mineral ((Mg, Fe)SiO₃), or to a perovskite-dominated assemblage (refs 1–4). Because silicate perovskite is stable to at least 70 GPa, it is thought to be the most abundant mineral in the lower mantle and, possibly, in the entire Earth. Despite its importance, silicate perovskite was only discovered in 1976, and little is known about its physical properties because of the difficulty in achieving the conditions of pressure and temperature required for the synthesis of this phase. For example, the value of the thermal expansion coefficient of perovskite provides an important constraint on possible compositional models for the lower mantle^{5–7}. We have recently produced sufficient amounts of (Mg_{0.9}, Fe_{0.1})SiO₃ perovskite to measure its zero-pressure thermal expansion to 840 K by X-ray diffraction. At high temperatures, the average thermal expansion coefficient is $4 \times 10^{-5} \text{ K}^{-1}$. Such a large value for the thermal expansion coefficient implies that standard models of upper mantle composition, such as pyrolite or garnet peridotite (Mg value ≈ 0.89), yield zero-pressure densities that are about 2% lower than that of the density of the lower mantle extrapolated to zero pressure conditions. This result suggests that the upper and lower mantle are chemically distinct, in accord with layered models of the thermal and convective state of the mantle.

We have measured the thermal expansion coefficient of (Mg_{0.9}, Fe_{0.1})SiO₃ perovskite using a high-temperature Guinier X-ray diffraction camera, and diffraction patterns at room and higher temperature were collected on film. The perovskite was synthesized from a natural orthopyroxene ((Mg_{0.88}, Fe_{0.12})SiO₃) above 35 GPa (350 kbar) and 2,000 K using laser heating through a Mao–Bell type diamond-anvil cell⁸. No pressure medium was used in the synthesis, and no ruby was used for pressure calibration, to prevent contamination of the perovskite. The sample pressure for the synthesis runs was determined from the force applied to the diamonds, which was independently calibrated as a function of pressure by the ruby fluorescence technique^{9,10}. Lattice parameters of the quenched perovskite at ambient conditions were determined from the (002), (110), (111), (020), (112), (200), (210) and (211) diffraction lines, and gave lattice parameters of $a = 478.3 (\pm 0.7) \text{ pm}$, $b = 495.3 (\pm 0.8) \text{ pm}$, $c = 688.6 (\pm 0.7) \text{ pm}$ and a density $\rho = 4.216 (\pm 0.004) \text{ Mg m}^{-3}$, in good agreement with previous results^{11–13}. The high-temperature experiments were performed on quenched perovskite samples at zero pressure to temperatures exceeding 900 K (630 °C). Each high-temperature run required 35–50 diamond cell samples, which were loaded into a 200- μm fused-quartz capillary tube. NaCl was mixed with the perovskite to act as a temperature calibrant. At high temperatures, the lattice parameters and, therefore, the high-temperature volumes were determined from the (110), (020), (112), (200) and (211) diffraction lines. The perovskite, which is metastable at room temperature and pressure, reverted to enstatite between 840 and 875 K (567–602 °C) as evidenced from a rapid change in the X-ray diffraction pattern between these temperatures. Figure 1 shows the increase in volume (decrease in density) with temperature for perovskite at

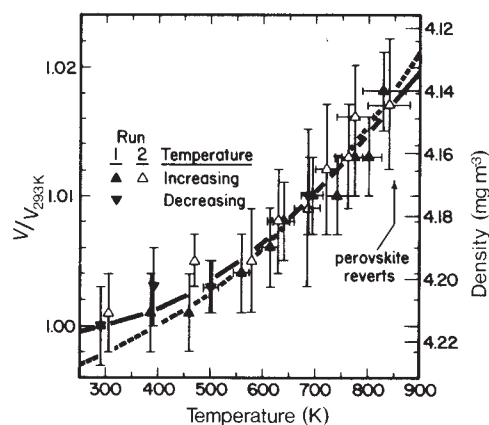


Fig. 1 The volume expansion of (Mg_{0.9}Fe_{0.1})SiO₃ perovskite at zero pressure as measured by X-ray diffraction from room temperature to 840 K. Above 840 K the high-pressure perovskite structure reverts to the low-pressure structure of orthopyroxene. The curves through the data represent two fits to the measurements using the Suzuki formalism for Gruneisen's theory of thermal expansion. The solid line is the best fit to all of the data and is used to obtain the thermal expansion coefficient from 0 to 2,500 K. The dashed line is a fit to the high-temperature data only. Note that in the high-temperature region, between 450 and 850 K, the volume increases by 1.6%, which corresponds to an average thermal expansion coefficient of $4 \times 10^{-5} \text{ K}^{-1}$.

zero pressure. The b/a and c/a lattice-parameter ratios of the perovskite remained constant within the error bars to 840 K indicating no further distortion in the perovskite structure at high temperature. Similarly, Yagi *et al.*¹⁴ found no conclusive evidence for a change in structural distortion with hydrostatic pressure to 8.4 GPa. From these two lines of evidence, we conclude that silicate perovskite does not undergo significant distortion from its zero-pressure orthorhombic structure at high pressure and temperature. Therefore, the physical properties of the quenched perovskite measured in the laboratory are relevant to silicate perovskite which exists at high pressure and temperature in the Earth's mantle.

Because the thermal expansion of perovskite cannot be measured above 840 K at zero pressure, our data must be extrapolated to temperatures for the Earth's mantle. We calculated the high-temperature thermal expansion coefficients using Gruneisen's theory of thermal expansion as formulated by Suzuki *et al.*^{15,16}. In this procedure, we fit the measured data to obtain values for the Gruneisen parameter γ , the Debye temperature θ , the pressure derivative of the bulk modulus K'_0 and the zero-temperature molar volume for perovskite (Table 1). Two acceptable fits to the data were obtained as indicated in Fig. 1.

Comparison of the high-temperature, zero-pressure perovskite densities can be made with the density and temperature of the decompressed lower mantle. Based on seismic observations, the bulk of the lower mantle appears to be homogeneous and adiabatic^{17–19}, and an Eulerian finite strain formalism can be used to extrapolate the density of the lower mantle adiabatically to zero-pressure, high-temperature conditions as was first done by Birch^{19–22}. These calculations give a decompressed value of $4.00 \pm 0.01 \text{ Mg m}^{-3}$ for the zero-pressure, high-temperature density of the high-pressure mineral assemblage comprising the lower mantle.

The temperature of the lower mantle on adiabatic decompression is uncertain. However, an estimated lower bound for this temperature, 1,700 (± 250) K ($\sim 1,400$ °C), is based on the phase equilibria of the olivine-spinel phase transitions which are thought to occur at 400 km depth^{20,23–25}. If the mantle is stratified and the 670-km discontinuity is a thermal as well as a chemical boundary layer, the zero-pressure temperature of the lower mantle adiabat may exceed 2,000 K (ref. 25).

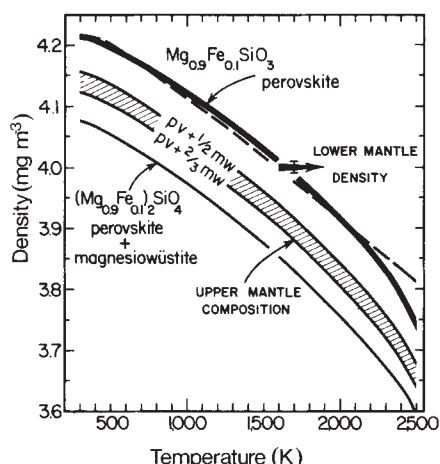


Fig. 2 Densities of proposed lower mantle assemblages (curves) are compared with the adiabatically decompressed values of density and temperature for the lower mantle (bold arrow). The zero-pressure value of the lower mantle density is well constrained ($4.00 \pm 0.01 \text{ Mg m}^{-3}$), but only a minimum value for the lower mantle adiabat (zero-pressure temperature of 1,700 K) is known with confidence: the bold arrow illustrates that higher temperatures are possible. The zero-pressure density of $(\text{Mg}, \text{Fe})\text{SiO}_3$ perovskite was extrapolated to high temperatures using the thermal expansion coefficient measured in this study (the solid and dashed bold curves correspond to the best fit and the high-temperature fit, respectively, shown in Fig. 1). A mineralogical model that satisfies the observed lower-mantle density consists of $(\text{Mg}, \text{Fe})\text{SiO}_3$ pyroxene alone, in its high-pressure perovskite phase. Lower-mantle assemblages consistent with a uniform mantle composition are indicated by the shaded band labelled 'upper mantle composition' (bounded by compositions between 1:1 and 2:1 olivine:pyroxene ratios, with pv and mw referring to perovskite and magnesiowüstite phases). All the models are based on $X_{\text{Mg}} = 0.89$ (refs 27, 28) and the measured pv-mw partition coefficients^{29,30}. None of the proposed upper mantle compositions or olivine alone (1:1 pv:mw) satisfy the observed lower mantle density at the appropriate conditions.

Figure 2 shows the effect of temperature on the densities of perovskite (pv) and several perovskite-magnesiowüstite (pv-mw) assemblages compared at zero pressure with the density and temperature inferred for the lower mantle. The thermal expansion of perovskite is extrapolated to high temperatures using the data summarized in Table 1, whereas that of magnesiowüstite has been measured to 1,300 K (the mean thermal expansion coefficient from 300 to 1,300 K is $4.8 \times 10^{-5} \text{ K}^{-1}$)^{15,26}. The region labelled upper mantle composition on Fig. 2 is bounded by the high-pressure phases which would result from a composition of olivine:pyroxene ratio between 1:1 and 2:1 (refs 27, 28). In each case, the upper mantle minerals are assumed to contain a ratio of magnesium to iron plus magnesium endmember components (X_{Mg}) of 0.89. The partitioning of iron between the resulting high-pressure phases was calculated using the partition coefficients determined by Bell *et al.*²⁹ and Ito and Yamada³⁰. Also included in Fig. 2 is the change in density with temperature for an olivine-only composition (1 pv:1 mw). It can be seen that the olivine-rich representative upper mantle compositions do not coincide with current estimates of the zero-pressure density and minimum adiabat temperature of the lower mantle (bold arrow).

To match the observed density of the lower mantle, a small but apparently significant enrichment of iron or silica (or both) is required relative to the composition of the upper mantle. If only iron enrichment is considered (upper mantle olivine:pyroxene ratio of 1:1 to 2:1) a value $X_{\text{Mg}} \leq 0.84 (\pm 0.02)$ is required to match the zero-pressure density of the lower mantle ($4.00 (\pm 0.01) \text{ Mg m}^{-3}$ at 1,700 K). In contrast, current estimates of the average upper mantle composition yield $X_{\text{Mg}} = 0.89$

Table 1 Thermal expansion parameters for $(\text{Mg}, \text{Fe})\text{SiO}_3$ perovskite

	γ	$\Theta(\text{K})$	K'_0	Molar volume at 0 K ($V(0 \text{ K})$, $\text{cm}^3 \text{ mol}^{-1}$)
Best fit	1.77	525	7	24.46
High-temperature fit	2.20	825	4	24.39

Equation for volume at temperature T (see refs 15, 16):

$$V(T) = \frac{V(0 \text{ K})}{2k} \left[1 + 2k - ((1 - 4kE)/Q)^{1/2} \right]$$

where $k = (K_0 - 1)/2$; $E(T, \Theta) = 45 RT(T/\Theta)^{3\Theta/T} z^3 / (e^z - 1) dz$; $Q = K_0 V(0 \text{ K})/\gamma$; $K_0 = 260 (\pm 20) \text{ GPa}$ (ref. 14).

(± 0.02) (refs 6, 7, 27, 28). Our results are unlikely to be sensitive to the presence of minor components such as aluminium or calcium^{7,20}. Nevertheless, we cannot uniquely determine the composition that is required to satisfy the observed properties of the lower mantle because of the tradeoff between the effects of iron content, silica content and temperature. For example, a pure perovskite composition with $X_{\text{Mg}} = 0.87 (\pm 0.01)$ coincides with the lower mantle density at 2,000 K, an adiabat temperature consistent with the presence of a thermal boundary layer associated with a compositional change at 670-km depth²⁵.

The idea that the upper and lower mantle are chemically distinct has been previously suggested from geophysical, petrological and geochemical arguments³¹⁻³⁴. Based on our new data on the high-temperature properties of perovskite we also find that the conventional model of upper mantle composition (such as peridotite with $X_{\text{Mg}} = 0.89$ or pyrolite^{27,28} seems not to be consistent with the observed lower mantle density. We cannot distinguish between possible iron enrichment or silica enrichment of the lower mantle, but are led to conclude that the 670-km discontinuity represents a change in average chemical composition of the Earth's mantle. Compositional stratification at this depth, in turn, suggests that the convection pattern within the mantle is predominantly two-layered, and that the timescale for the thermal evolution of the Earth's interior is considerably longer than the 1,000–2,000 Myr expected with unlayered, whole-mantle convection³⁵. The presence of chemical stratification in the mantle depends on a delicate balance between the compositionally-induced density change and the Rayleigh number for lower mantle convection, because at sufficiently large Rayleigh number, convection would destroy any chemical layering (see refs 36–38). Although some leakage may occur between the upper and lower mantle^{19,39}, our data suggest that the intrinsically higher density of the lower (relative to the upper) mantle impedes mantle-wide mixing and results in a chemical and thermal boundary layer at 670-km depth.

We thank M. H. Manghnani, S. Akimoto and T. J. Ahrens for helpful comments. Work supported by the University of California Institute for Geophysics and Planetary Physics (L.L.N.L. branch).

Received 22 April; accepted 24 October 1985.

1. Akaogi, M. & Akimoto, S. *Phys. Earth planet. Inter.* **15**, 90–106 (1977).
2. Akimoto, S., Yagi, T. & Inoue, K. in *High-Pressure Research* (eds Manghnani, M. H. & Akimoto, S.) 585–692 (Academic, New York, 1977).
3. Liu, L. in *The Earth: Its Origin Structure and Evolution* (ed. McElhinny, M. W.) 177–202 (Academic, New York, 1979).
4. Yagi, T., Bell, P. M. & Mao, H.-K. *Yb. Carnegie Instn Wash.* **78**, 614–615 (1979).
5. Liu, L. *Earth planet. Sci. Lett.* **42**, 202–208 (1979).
6. Jackson, I. *Earth planet. Sci. Lett.* **62**, 91–103 (1983).
7. Jeanloz, R. & Thompson, A. B. *Rev. Geophys. Space Phys.* **21**, 51–74 (1983).
8. Mao, H.-K., Bell, P. M., Dunn, K. J., Chrenko, R. M. & Devries, R. C. *Rev. Scient. Instrum.* **50**, 1002–1009 (1979).
9. Barnett, J. D., Block, S. & Piermarini, G. J. *Rev. Scient. Instrum.* **44**, 1–9 (1973).
10. Mao, H.-K., Bell, P. M., Shaner, J. W. & Steinberg, D. J. *J. appl. Phys.* **49**, 3276–3283 (1978).
11. Liu, L. *Phys. Earth planet. Inter.* **11**, 289–298 (1976).
12. Ito, E. & Matsui, Y. *Earth planet. Sci. Lett.* **38**, 435–443 (1978).
13. Yagi, T., Mao, H.-K. & Bell, P. M. *Phys. Chem. Miner.* **3**, 97–103 (1978).

14. Yagi, T., Mao, H.-K. & Bell, P. M. *Physical Geochemistry* (ed. Saxena, S. K.) 317-325 (Springer, New York, 1982).
15. Suzuki, I. *J. Phys. Earth* **23**, 145-159 (1975).
16. Suzuki, I., Okajima, S. & Seya, K. *J. Phys. Earth* **27**, 63-69 (1979).
17. Masters, G. *Geophys. J. R. astr. Soc.* **57**, 507-534 (1979).
18. Dziewonski, A. M. & Anderson, D. L. *Phys. Earth planet. Inter.* **25**, 297-357 (1981).
19. Birch, F. *J. geophys. Res.* **69**, 4377-4388 (1964).
20. Jeanloz, R. in *Convection in the Earth's Mantle* (ed. Peltier, W. R.) (Gordon and Breach, New York, in the press).
21. Davies, G. F. & Dziewonski, A. M. *Phys. Earth planet. Inter.* **10**, 336-343 (1975).
22. Dziewonski, A. M., Hales, A. L. & Lapwood, E. R. *Phys. Earth planet. Inter.* **10**, 12-48 (1975).
23. Akimoto, S., Akaogi, M., Kawada, K. & Nishizawa, O. in *The Geophysics of the Pacific Ocean Basin and Its Margin* (eds Sutton, G. H., Manghnani, M. H. & Moberly, R.) 399-405 (AGU, Washington DC, 1976).
24. Anderson, D. L. *Science* **157**, 1165-1173 (1967).
25. Jeanloz, R. & Richter, F. M. *J. geophys. Res.* **84**, 5497-5504 (1979).
26. Touloukian, Y. S., Kirby, R. K., Taylor, R. E. & Lee, T. Y. R. in *Thermal Expansion, Non-metallic Solids* (TPRC 13) 1658 (IFI/Plenum, New York, 1977).
27. Ringwood, A. E. *Composition and Petrology of the Earth's Mantle* (McGraw-Hill, New York, 1975).
28. Yoder, H. S. Jr *Generation of Basaltic Magmas* (National Academy of Sciences, Washington DC, 1976).
29. Bell, P. M., Yagi, T. & Mao, H.-K. *Yb. Carnegie Instn. Wash.* **78**, 618-619 (1979).
30. Ito, E. & Yamada, H. in *High Pressure Research in Geophysics* (eds Akimoto, S. & Manghnani, M. H.) 405-419 (Center for Academic Publishing, Tokyo, 1982).
31. Lees, A. C., Bukowski, M. S. T. & Jeanloz, R. *J. geophys. Res.* **88**, 8145-8159 (1983).
32. Anderson, D. L. *A. Rev. Earth planet. Sci.* **5**, 179-202 (1977).
33. O'Nions, R. K., Eversen, N. M. & Hamilton, P. J. *J. geophys. Res.* **84**, 6091-6101 (1979).
34. Jacobsen, S. B. & Wasserburg, G. J. *J. geophys. Res.* **84**, 7411-7427 (1979).
35. McKenzie, D. & Richter, F. M. *J. geophys. Res.* **86**, 11667-11680 (1981).
36. Richter, F. M. & Johnson, C. E. *J. geophys. Res.* **79**, 1635-1639 (1974).
37. Olson, P. *J. geophys. Res.* **89**, 11293-11301 (1984).
38. Hoffman, N. R. A. & McKenzie, D. P. *Geophys. J. R. astr. Soc.* **82**, 163-206 (1985).
39. Creager, K. C. & Jordan, T. H. *J. geophys. Res.* **89**, 3031-3050 (1984).

$^{187}\text{Os}/^{186}\text{Os}$ in marine manganese nodules and the constraints on the crustal geochemistries of rhenium and osmium

Martin R. Palmer* & Karl K. Turekian

Department of Geology and Geophysics, Yale University, Box 6666, New Haven, Connecticut 06511, USA

During their formation, marine manganese (or more properly ferromanganese oxide) nodules include and commonly concentrate elements dissolved in seawater. This attribute makes them especially valuable recorders of the isotopic composition of elements supplied to the oceans having radiogenic isotopes, for these can be used in exploring crustal composition and evolution¹⁻⁵. From measurements of $^{187}\text{Os}/^{186}\text{Os}$ ratios in marine manganese nodules and comparison with the $^{143}\text{Nd}/^{144}\text{Nd}$ ratios of the same nodules, we have calculated the Os isotopic composition and the osmium and rhenium concentrations of the weatherable portion of the continental crust, together with the proportions of 'ultramafic' and 'granitic' rock that constitute this crustal mass.

Our earlier studies in this laboratory^{5,6} sought to relate the Os isotopic composition of marine manganese nodules to the sources of Os in the oceans. ^{187}Re decays ($\lambda = 1.52 \times 10^{-11} \text{ yr}^{-1}$) to form ^{187}Os . Variations in the ^{187}Os content of Os are commonly reported as the $^{187}\text{Os}/^{186}\text{Os}$ ratio since ^{186}Os is not radiogenic and is approximately equal in abundance to ^{187}Os in primitive materials such as meteorites⁷. The $^{187}\text{Os}/^{186}\text{Os}$ ratio depends on the age of the system and its $^{187}\text{Re}/^{186}\text{Os}$ ratio. The $^{187}\text{Os}/^{186}\text{Os}$ in manganese nodules records the mixing in the oceans of Os from systems with different ages and $^{187}\text{Re}/^{186}\text{Os}$.

To unravel these sources and their properties, additional information is required. $^{87}\text{Sr}/^{86}\text{Sr}$ is a possible aid, but it depends mainly on temporal changes rather than regional diversity to recover such information⁵. Lead isotopes are complex and show perhaps too high a sensitivity to their sources¹ for our purposes. Variations in the $^{143}\text{Nd}/^{144}\text{Nd}$ ratio in manganese nodules have been well established and appear to provide the necessary constraint to the understanding of the $^{187}\text{Os}/^{186}\text{Os}$ texture of the

oceans. Turekian and Luck⁶ made an initial attempt at coupling the $^{187}\text{Os}/^{186}\text{Os}$ and $^{143}\text{Nd}/^{144}\text{Nd}$ systems to decipher sources of Os to the oceans. We now report additional Os isotopic data on nodules previously analysed for Nd isotopes and refine the interpretations made earlier.

The experimental procedures for measuring Os isotope ratios in manganese nodules have been discussed elsewhere^{5,8}. Briefly, the technique consists of dissolving 10 g of the sample in a warm oxalic acid solution followed by distillation to extract the volatile OsO_4 . This is then converted to OsCl_6^{2-} , purified on an ion-exchange column and plated on a silicon chip. The chloride salt is converted to the metal in a vacuum furnace. The actual Os isotopic composition is determined by secondary-ion microprobe mass spectrometry using the MIT-Brown-Harvard consortium Cameca IMS-3f ion microprobe. Osmium concentrations were determined in a few samples by isotope dilution.

The $^{187}\text{Os}/^{186}\text{Os}$ ratios measured in this study are listed in Table 1, together with those determined by Luck and Turekian⁵. All errors are 2σ of the mean. The data set is too small to provide an unambiguous judgement about details of the regional differences in $^{187}\text{Os}/^{186}\text{Os}$, but the following observations can be made: (1) the mean values of the Atlantic, Pacific and Indian oceans are indistinguishable; (2) the largest range in $^{187}\text{Os}/^{186}\text{Os}$ of these basins occurs in the Atlantic (5.2-10.9); (3) some areas of an ocean basin have relatively homogeneous values, whereas others (such as, the eastern North Atlantic) show large differences within relatively small distances (see Fig. 1).

Turekian and Luck⁶ have demonstrated that the Os extracted from manganese nodules, using the techniques discussed above, is virtually free of ^{187}Os produced *in situ* from the decay of ^{187}Re or Os associated with meteoritic debris trapped in the nodule. Although they addressed the problem of the contribution of magnetic spherules of meteoritic origin included in nodules, they assumed that all meteoritic debris is deposited on the ocean floor as particles. However, if all the meteoritic debris did go into solution, the meteoritic flux of osmium to seawater could be three times the riverine flux⁹. This would virtually homogenize the $^{187}\text{Os}/^{186}\text{Os}$ of the oceans. Clearly, this is not the case. Brownlee *et al.*¹⁰ have recently demonstrated that Os and Ir occur as metallic 'nuggets' of $\sim 10\text{-}\mu\text{m}$ diameter within magnetic spherules produced by the oxidation of meteoritic iron grains, probably during passage through the atmosphere. Therefore, because Os and Ir remain as refractory metals, they are less likely to be dissolved than some other elements of meteoritic origin. Hence, the measured $^{187}\text{Os}/^{186}\text{Os}$ ratios are the result of the integration of sources of dissolved Os with different isotopic compositions supplied to the oceans, primarily from terrestrial sources, and deposited in the nodule.

The strong differences in $^{187}\text{Os}/^{186}\text{Os}$ within ocean basins, and in some cases between nodules in close proximity, indicate that Os has a short mean oceanic residence time ($\sim 10^2$ yr). For this reason, the relative input rates of Os of different isotopic compositions to each basin will be imprinted in the manganese nodules forming there.

We now characterize the potential sources of Os to the major ocean basins by coupling measurements of $^{143}\text{Nd}/^{144}\text{Nd}$ and $^{187}\text{Os}/^{186}\text{Os}$ in manganese nodules. The result of these calculations yields the $^{187}\text{Os}/^{186}\text{Os}$ ratio of the continental crust. This will provide constraints on the average Re/Os ratio of the crust and point to directions for verification. It will also provide information concerning the sources of mantle-derived Os to the different ocean basins. We draw on a model previously described⁶ but with some modifications. Whereas before regional $^{143}\text{Nd}/^{144}\text{Nd}$ values were used, in the present study both Os and Nd isotopic measurements are available for each nodule.

We treat each of the major oceans separately, starting with the Pacific because of its size, relatively simple water structure compared with the Atlantic and the greater homogeneity of the $^{187}\text{Os}/^{186}\text{Os}$ ratios there.

To determine the $^{187}\text{Os}/^{186}\text{Os}$ ratio of the continental crust

* Present address: Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

Table 1 Os concentration, $^{187}\text{Os}/^{186}\text{Os}$ and $\epsilon_{\text{Nd}}(\text{O})$ of marine manganese nodules

Sample	Location		[Os] (p.p.b.)	$^{187}\text{Os}/^{186}\text{Os}$		$\epsilon_{\text{Nd}}(\text{O})^\dagger$
Atlantic						
V19 D11	38°04' N	60°13' W		8.09 ± 0.06*		(−12.0)
V27 D11	42°20' N	27°0' W		7.97 ± 0.11*		(−11.6)
V27 133	45°05' N	7°56' W		10.96 ± 0.40		(−11.5)
VEMA FZ	10°54' N	45°18' W	2.39 ± 0.02	8.11 ± 0.18	8.01 ± 0.08	−10.6
RC 16 D10	28°25' S	40°45' W		6.68 ± 0.10*		(−14.0)
V16 SBT3	13°04' S	24°41' W		7.44 ± 0.20		−10.9
V26 57	20°46' S	24°08' W		6.66 ± 0.19		(−10.8)
V15 138	49°35' S	48°05' W		7.56 ± 0.26		−8.4
CHN115 146	30°13' S	39°4' W	1.59 ± 0.13	7.68 ± 0.19	7.23 ± 0.20	−14.0
V27 D5	56° N	21° W		5.27 ± 0.19	5.13 ± 0.22	−10.3
BP 2381	31°4' N	78°9' W		8.49 ± 0.13		−11.0
KNR42-166	35°36' N	58°41' W		8.30 ± 0.09	8.41 ± 0.07	−11.4
V27 58	75°32' N	2°40' E		7.03 ± 0.18		−10.0
Indian						
RC14 D3	34°45' S	44°46' E	0.82 ± 0.09	8.00 ± 0.19	7.39 ± 0.07	−9.7
RC14 D5	19°2' S	61°1' E		8.33 ± 0.09*		(−8.4)
DODO 232D	5°23' S	97°29' E	0.84 ± 0.02	7.02 ± 0.18	7.20 ± 0.11	−7.3
DODO 127D	6°40' S	51°54' E		8.10 ± 0.17		−7.8
DODO 62D	16°18' S	104°16' E		7.37 ± 0.16		−7.3
Pacific						
MN 139	20°01' N	136°36' E	0.87 ± 0.02	8.27 ± 0.15	8.07 ± 0.08	−3.6
A47 16	9°02' N	151°11' W		7.65 ± 0.07*		(−3.9)
Challenger 276	13°28' S	149°30' W		5.95 ± 0.18*		(−3.4)
A50 D	32°42' N	158°15' E		8.91 ± 0.26		−3.7
V28 D8	0° N	179° W		7.50 ± 0.26		−3.4
Antp 580	18°57' N	135°48' W	1.95 ± 0.07	7.80 ± 0.12	7.84 ± 0.06	−4.3
E11 USC 887	56°54' S	115°13' W		8.38 ± 0.19*		(−6.6)
SCAN 35D	20°55' N	142°22' E		8.23 ± 0.09		−3.5
MN 171 DJ25	9°22' N	151°12' W		7.00 ± 0.16		(−3.9)
MN 196 DJ52	9°20' N	151°27' W		7.20 ± 0.22		(−3.9)
MN 203 DJ59	9°18' N	151°32' W		7.06 ± 0.31		(−3.9)

$\epsilon_{\text{Nd}}(\text{O})$ value in parentheses indicates that a value is not available for that nodule and Nd isotopic data from the nearest available nodule are given.

* From ref. 5; † refs 2-4.

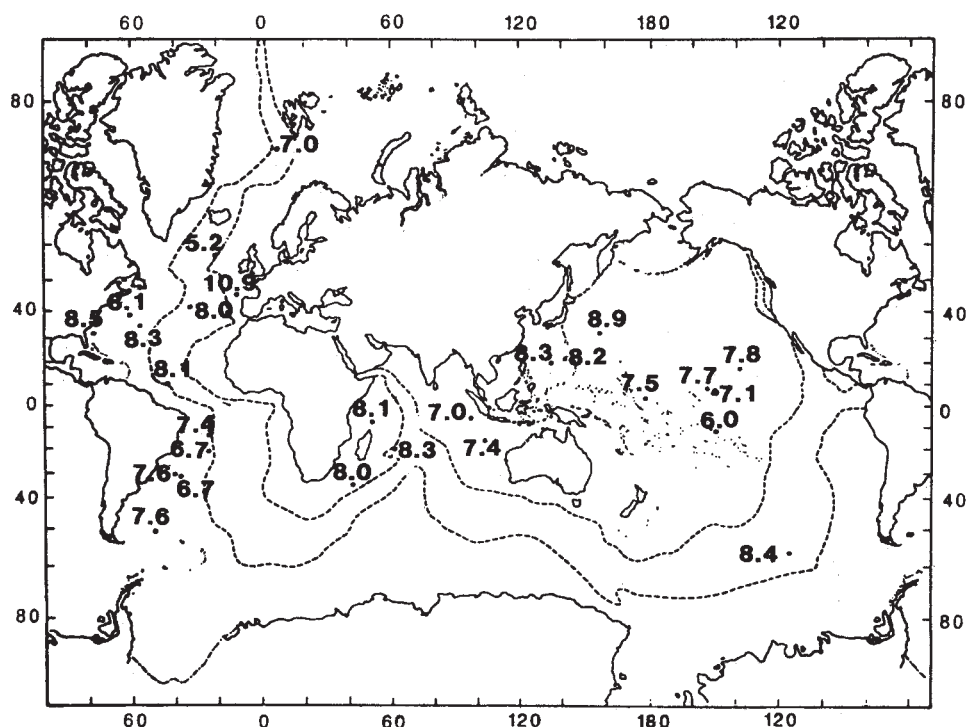
Fig. 1 $^{187}\text{Os}/^{186}\text{Os}$ ratios in marine manganese nodules.

Table 2 Ir (Os) and Nd concentrations and $^{187}\text{Os}/^{186}\text{Os}$ and $\epsilon_{\text{Nd}}(\text{O})$ values of sources to the oceans

Source	[Ir] (p.p.b.)*	[Nd] (p.p.m.)	[Ir]/[Nd]	$^{187}\text{Os}/^{186}\text{Os}$	$\epsilon_{\text{Nd}}(\text{O})$
Continental crust					
Precambrian shield composite ²⁶	0.05	19.4	2.58×10^{-6}	—	—
Mississippi suspended particulates ²⁷	0.056	28.9	1.94×10^{-6}	—	—
Mississippi sediment† (corrected for authigenic component) ^{12,27}	0.081	32.6	2.48×10^{-6}	—	—
Ocean island basalt ^{28-30†}	0.44	35	12.6×10^{-6}	1	+6
Peridotite ^{7,20†}	5	0.5	10.6×10^{-3}	1	+10

* [Ir] ≡ [Os].

† Values used for calculations in text.

that is recorded by the Pacific nodules, we make the following assumptions, most of which are reasonable in the face of our current ignorance: (1) the present-day $^{187}\text{Os}/^{186}\text{Os}$ ratio of mantle-derived material is 1 (ref. 7); (2) the dominant source of mantle-derived Nd and Os in the Pacific is the alteration (low temperature and hydrothermal) of ocean island basalt (OIB) represented by Hawaii and other 'hotspot' islands and seamounts; (3) the supply of Os and Nd from tholeiitic mid-ocean ridge hydrothermal vent systems is negligible¹¹; (4) Nd and Os are released congruently during weathering and alteration; (5) the continental $^{143}\text{Nd}/^{144}\text{Nd}$ ratio supplied to the Pacific Ocean is the average of values for the major rivers flowing into the Pacific Ocean basin¹²; (6) the mean oceanic residence times of Os and Nd are short relative to the mixing times of the oceans (see above); (7) Nd and Os in labile sites in deep sea deposits define the isotopic composition of provinces within the ocean basins; (8) in the absence of a significant amount of reliable crustal Os concentration data, we assume the Os/Ir ratio is approximately 1, as determined in meteorites and ultramafic rocks¹³⁻¹⁵. This ratio refers to the component of Os having a mantle isotopic imprint.

The continental crustal $^{187}\text{Os}/^{186}\text{Os}$ ratio can then be obtained from⁶:

$$R_c = \frac{\frac{[\text{Os}]_c}{[\text{Nd}]_c} \frac{\epsilon_b - \epsilon_c}{\epsilon_s - \epsilon_c} R_s + \frac{[\text{Os}]_b}{[\text{Nd}]_b} (R_s - R_b)}{\frac{[\text{Os}]_c}{[\text{Nd}]_c} \times \frac{\epsilon_b - \epsilon_s}{\epsilon_s - \epsilon_c}}$$

where R_c is the $^{187}\text{Os}/^{186}\text{Os}$ ratio of the continental crust; R_s is the $^{187}\text{Os}/^{186}\text{Os}$ ratio of the sample; $R_b = 1$, $^{187}\text{Os}/^{186}\text{Os}$ ratio of the 'mantle' source (OIB); $\epsilon_c = -10.7$, $\epsilon_{\text{Nd}}(\text{O})$ of the continental crust supplying the Pacific Ocean; ϵ_s is the $\epsilon_{\text{Nd}}(\text{O})$ of the sample; $\epsilon_b = +6$, $\epsilon_{\text{Nd}}(\text{O})$ of the 'mantle' source (OIB); $[\text{Nd}]_c = 32.6$, Nd concentration (p.p.m., parts per 10^6) of the crust; $[\text{Nd}]_b = 35$, Nd concentration (p.p.m.) of the 'mantle' source (OIB); $[\text{Os}]_c = 0.081$, the Os concentration (p.p.b., parts per 10^9) of the crust; $[\text{Os}]_b = 0.44$, the Os concentration (p.p.b.) of the 'mantle' source (OIB).

The sources of the data are listed in Table 2. Note that the absolute concentrations of Os and Nd are not required, but rather their ratios in the major rock types are used in the calculations. In crustal rocks, the Ir/Nd ratio is fairly constant (see Table 2). A sensitivity test conducted for reasonable variations in the parameters used in the equation above demonstrated that the calculated continental crustal $^{187}\text{Os}/^{186}\text{Os}$ ratio does not vary enough to alter significantly the conclusions presented below. (Hydrothermal ferromanganese deposits are not considered here.)

This calculation has been performed for all the Pacific nodules and the results are summarized in Table 3. The average $^{187}\text{Os}/^{186}\text{Os}$ ratio for the continental crust is 30. Thus ~25% of the Os found in Pacific manganese nodules is from continental

Table 3 $^{187}\text{Os}/^{186}\text{Os}$ of continental crust

Sample	$\epsilon_{\text{Nd}}(\text{O})$ nodule*	$^{187}\text{Os}/^{186}\text{Os}$ Nodule	Continental crust
MN 139	-3.6	8.27	35.5
A47 16	(-3.9)	7.65	30.7
Challenger 276	(-3.4)	5.95	25.4
A50 D	-3.7	8.91	37.8
V28 D8	-3.4	7.50	33.0
Antp 580	-4.3	7.80	29.2
E11 USC 887	(-6.6)	8.38	20.5
SCAN 35D	-3.5	8.23	35.9
MN 171 DJ25	(-3.9)	7.00	27.8
MN 196 DJ52	(-3.9)	7.06	28.1
MN 203 DJ59	(-3.9)	7.20	28.7
Mean			30.2

$\epsilon_{\text{Nd}}(\text{O})$ value in parenthesis indicates that a value is not available for that nodule and Nd isotopic data from the nearest available nodule is given.

* From refs 2-4.

crustal sources and the remaining 75% is from OIB alteration.

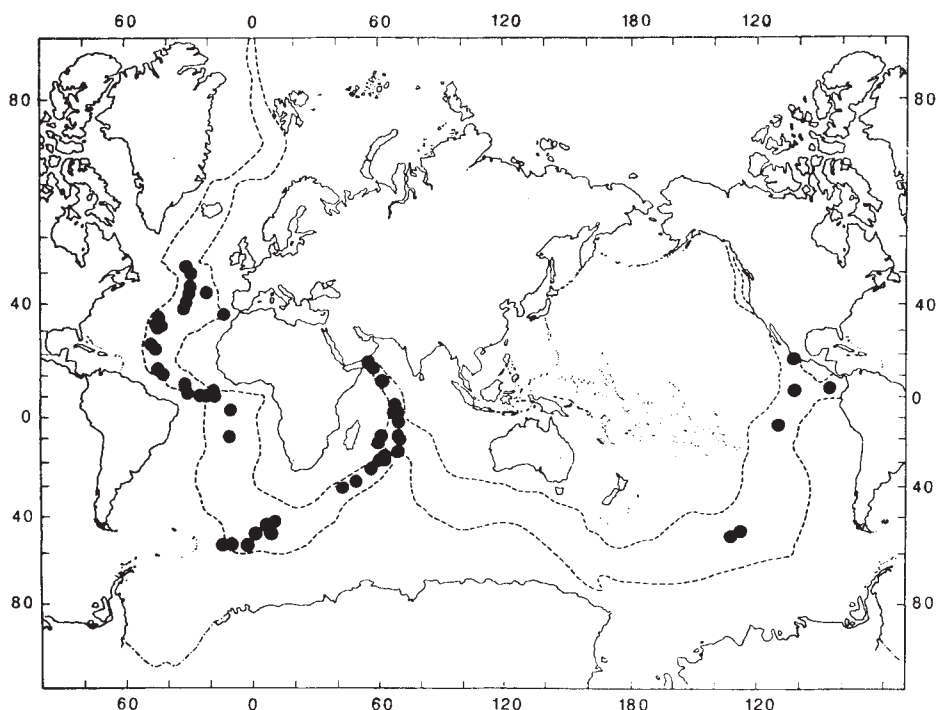
If we assume that the average age of the continental crust being drained by the major rivers flowing into the Pacific is equal to that of the continental crust being drained by Atlantic rivers¹², ~25% of the Os in the Atlantic manganese nodules is also derived from the continental crust, because the mean $^{187}\text{Os}/^{186}\text{Os}$ of both oceans is approximately the same. (Actually the average $^{143}\text{Nd}/^{144}\text{Nd}$ of the Atlantic feeding rivers is slightly lower than those for the Pacific¹², but the differences are too small to be of significance.) Atlantic manganese nodules, however, unlike those from the Pacific, have $\epsilon_{\text{Nd}}(\text{O})$ values that fall within the range for the average continental crustal value of -11.4 ± 4.0 (ref. 12). The source of the mantle-derived Os cannot, therefore, be the alteration of high Nd rocks such as OIB that are so abundant in the Pacific. We believe that these high Os, low Nd sources in the Atlantic are ultramafic rocks. Specifically, we identify them as the commonly serpentinized ultramafic rocks associated with the fracture zones along the mid-Atlantic ridge.

The Os concentration of peridotites is approximately two orders of magnitude higher than that of the continental crust and one order higher than OIB. Because ocean floor peridotites are virtually all strongly altered to serpentinite¹⁶, the Os in peridotites is probably available for release to the oceans at the time of alteration. These aspects of the process have still to be investigated. Ocean floor peridotites are generally restricted to the walls of fracture zones associated, with the relatively slow spreading mid-Atlantic and the Atlantic-Indian ridges¹⁶⁻¹⁹ (see Fig. 2). Very few peridotites have been recovered from the Pacific fracture zones. The low concentration of Nd in peridotites²⁰, relative to other mantle oceanic rocks indicates that they are not a significant source of Nd during submarine weathering or hydrothermal alteration.

From the continental crustal $^{187}\text{Os}/^{186}\text{Os}$ ratio of 30 we can determine the average Re/Os ratio by assuming a mean continental crust age of 2×10^9 yr (ref. 12, for example) and the establishment of the crust 4×10^9 yr ago. Assuming a constant rate of continental growth and equal availability to weathering of all age rocks, a $^{187}\text{Os}/^{186}\text{Os}$ ratio would correspond to a $^{187}\text{Re}/^{186}\text{Os}$ of ~1000, that is, an elemental Re/Os ratio of ~25. Thus the continental crustal Os concentration (Table 2) yields a Re concentration range between 1.25 and 2 ng Re g⁻¹.

An independent estimate of the Re concentration of the continental crust can be obtained by using its close relationship to Mo in crustal rocks. This is most obviously seen in the high concentrations of Re in molybdenites²¹. Newsom and Palme²² have estimated that the Mo concentration of the upper mantle is $0.059 \mu\text{g Mo g}^{-1}$. Chou *et al.*¹⁵, using a slightly different

Fig. 2 Distribution of peridotites recovered from fracture zones at mid-ocean ridges¹⁶⁻¹⁹.



method arrive at a Re concentration in the upper mantle of $0.25 \text{ ng Re g}^{-1}$. This yields a Re/Mo ratio of 0.0042. Estimates of the Mo concentrations of the continental crust range between 1 and $1.5 \mu\text{g Mo g}^{-1}$ (see ref. 23 for a summary). An average value of $1.3 \mu\text{g Mo g}^{-1}$ gives a continental crustal Re concentration of 5.5 ng Re g^{-1} , assuming no fractionation between mantle and crust. If this is correct, then the Os concentration of the continental crust, based on a $^{187}\text{Os}/^{186}\text{Os}$ ratio of 30 and a Re/Os ratio of 25 (see above), would be $0.22 \text{ ng Os g}^{-1}$. This value is much higher than the value of 0.08 based on the Ir concentrations discussed above.

Assuming the Re/Mo ratio of the mantle determined from the data of Chou *et al.*¹⁵ and Newsom and Palme²² is correct, it is more likely that fractionation of Re and Mo during the formation of the crust from the mantle has resulted in a crustal Re/Mo ratio that is one-third that of the mantle.

Turekian and Luck⁶ used a simple model of weathering to explain the crustal $^{187}\text{Os}/^{186}\text{Os}$ ratio calculated for the Atlantic and Pacific basins. The difference between the Atlantic and Pacific Os and Nd isotopic signatures requires an ultramafic source to contribute Os more efficiently to the Atlantic Ocean than the Pacific Ocean. We have reaffirmed that the differences are due to the alteration of ultramafic rocks uniquely present within the Atlantic Ocean basin. We have assumed that all the continents yield the same $^{187}\text{Os}/^{186}\text{Os}$ ratio. Turekian and Luck⁶ allowed for the possibility that the 'ultramafic' to 'granite' ratio could vary between the continents and, therefore, the $^{187}\text{Os}/^{186}\text{Os}$ ratio of the continents could differ. This mechanism is not required if, as has been demonstrated here, the internal ocean basin imprints are significant and vary between the Atlantic and Pacific.

We can continue to use their model to calculate the relative proportions of 'ultramafic' and 'granitic' rocks that are undergoing weathering worldwide. The model partitions the continental crust into an 'ultramafic' component with $^{187}\text{Os}/^{186}\text{Os} = 1$ and a purely radiogenic ^{187}Os component associated with high Re-bearing mineral phases of 'granitic' rocks. These Re-bearing minerals are probably those rich in Mo, most notably molybdenite²¹.

The amounts of 'granitic' and 'ultramafic' rocks undergoing weathering are designated A_G and A_U , respectively; the continental crustal $^{187}\text{Os}/^{186}\text{Os} = 30$ and ultramafic rock

$^{187}\text{Os}/^{186}\text{Os} = 1$; the Os concentration of ultramafic rock is 5 ng Os g^{-1} (refs 7, 15); the mean age of the continents is $2 \times 10^9 \text{ yr}$; $[^{187}\text{Os}]_G = [^{187}\text{Re}]_G \lambda t$. With these parameters we can determine the relationship between the concentration of Re in the 'granitic' crust and the ratio of 'ultramafic' to 'granitic' rocks being weathered (A_U/A_G). For the 'ultramafic' contribution we get:

$$[^{187}\text{Os}]_U = 0.016[\text{Os}]_U = 0.016 \times 5 \text{ ng } ^{187}\text{Os g}^{-1} \\ = 0.08 \text{ ng } ^{187}\text{Os g}^{-1}$$

and for the 'granitic' contribution we get:

$$[^{187}\text{Os}]_G = 0.63[\text{Re}]_G \lambda t \\ = [\text{Re}]_G \times 0.63 \times 1.52 \times 10^{-11} \text{ yr}^{-1} \times 2 \times 10^9 \text{ yr}$$

(0.016 and 0.63 are the fractions of ^{186}Os and ^{186}Re in Os and Re respectively).

The relationship between $[\text{Re}]_G$ and A_U/A_G in these conditions (assuming total dissolution during weathering) is:

$$[\text{Re}]_G = 121.4 A_U/A_G$$

If $[\text{Re}]_G = 2.0$ then $A_U/A_G = 0.0165$, a value close to Daly's²⁴ estimate of the relative proportions of 'ultramafic' and 'granitic' rocks in the continental crust. To obtain A_U/A_G close to Knopf's²⁵ estimate (0.0025) requires $[\text{Re}]_G$ to be an order of magnitude lower than the value used here.

We thank J. M. Luck and N. Shimizu for technical help, D. J. Piepgras, S. L. Goldstein and S. E. Calvert for donating samples, and S. R. Hart for permission to use the SIMS ion probe at MIT. M. Lindner of the Lawrence Livermore National Laboratory made and provided us with ^{185}Os tracer. This research was funded by NSF grant OCE-8414753. M.R.P. acknowledges the partial support of a NATO fellowship. We thank R. K. O'Nions for comments.

Received 14 June; accepted 14 October 1985.

1. Chow, T. J. & Patterson, C. C. *Geochim. cosmochim. Acta* **26**, 263-308 (1962).
2. O'Nions, R. K., Carter, S. R., Cohen, R. S., Evenson, N. M. & Hamilton, P. J. *Nature* **273**, 435-438 (1978).
3. Piepgras, D. J., Wasserburg, G. J. & Dasch, E. J. *Earth planet. Sci. Lett.* **45**, 223-236 (1979).
4. Goldstein, S. L. & O'Nions, R. K. *Nature* **292**, 324-327 (1981).
5. Luck, J. M. & Turekian, K. K. *Science* **222**, 613-615 (1983).
6. Turekian, K. K. & Luck, J. M. *Proc. natn. Acad. Sci. U.S.A.* **81**, 8032-8034 (1984).
7. Allegre, C. J. & Luck, J. M. *Earth planet. Sci. Lett.* **48**, 148-154 (1980).

8. Luck, J. M. thesis, Univ. Paris 7 (1982).
9. Turekian, K. K. *Geol. Soc. Am. Spec. Pap.* **190**, 243–247 (1981).
10. Brownlee, D. E., Bates, B. A. & Wheelock, M. M. *Nature* **309**, 693–695 (1984).
11. Piepgras, D. J. & Wasserburg, G. J. *Earth planet. Sci. Lett.* **72**, 341–356 (1985).
12. Goldstein, S. L., O'Nions, R. K. & Hamilton, P. J. *Earth planet. Sci. Lett.* **70**, 221–236 (1984).
13. Barker, J. L. & Anders, E. *Geochim. cosmochim. Acta* **32**, 627–645 (1968).
14. Hertogen, J., Janssens, M. J. & Palme, H. *Geochim. cosmochim. Acta* **44**, 2125–2143 (1980).
15. Chou, C. L., Shaw, D. M. & Crockett, J. H. *J. geophys. Res.* Vol. 88 Suppl., A507–A518 (1983).
16. Bonatti, E. & Hamlyn, P. R. in *The Sea* Vol. 4 (ed. Emiliani, C.) (Wiley, New York, 1981).
17. Udintsev, G. B. & Dmitriev, L. V. in *The Sea* Vol. 4 (ed. Maxwell, A. E.) (Wiley, New York, 1971).
18. Hebert, R., Bideau, D. & Hekinian, R. *Earth planet. Sci. Lett.* **65**, 107–125 (1983).
19. Dick, H. J. W., Fisher, R. L. & Bryan, W. B. *Earth planet. Sci. Lett.* **69**, 88–106 (1984).
20. Frey, F. A. in *Rare Earth Element Geochemistry* (ed. Henderson, P.) (Elsevier, Amsterdam, 1984).
21. Luck, J. M. & Allègre, C. J. *Earth planet. Sci. Lett.* **61**, 291–296 (1982).
22. Newsom, H. E. & Palme, H. *Earth planet. Sci. Lett.* **69**, 354–364 (1984).
23. Manheim, F. T. in *Handbook of Geochemistry* II-4 (ed. Wedepohl, K. H.) (Springer, Berlin, 1969).
24. Daly, R. A. *Igneous Rocks and the Depths of the Earth* (McGraw-Hill, New York, 1933).
25. Knopf, A. J. *Geol.* **24**, 620–622 (1916).
26. Shaw, D. M., Dostal, J. & Keays, R. D. *Geochim. cosmochim. Acta* **40**, 73–83 (1976).
27. Fenner, F. D. & Presley, B. J. *Nature* **312**, 260–262 (1984).
28. White, W. H. & Hoffman, A. W. *Nature* **296**, 821–825 (1982).
29. Zoller, W. H., Parington, J. R. & Phelan Kotra, J. M. *Science* **222**, 1118–1121 (1983).
30. Sun, S. S. *Phil. Trans. R. Soc. Lond.* **297A**, 409–445 (1980).

Protein kinase C activation induces conductance changes in *Hermissenda* photoreceptors like those seen in associative learning

J. Farley* & S. Auerbach†

* Program in Neuroscience and Behavior, Department of Psychology, Princeton University, Princeton, New Jersey 08540, USA

† Department of Biological Sciences, Rutgers University, Piscataway, New Jersey 08854, USA

Phosphorylation of ion channels has been suggested as one molecular mechanism responsible for learning-produced long-term changes in neuronal excitability¹. Persistent training-produced changes in two distinct K⁺ currents (I_A (ref. 2), I_{K-Ca} (refs 3, 4)) and a voltage-dependent calcium current (I_{Ca} ; refs 3, 4) have previously been shown to occur in type B photoreceptors of *Hermissenda*, as a result of associative learning. But the identity of the phosphorylation pathway(s) responsible for these changes has not as yet been determined. Injections of cyclic AMP-dependent protein kinase⁵ reduce a K⁺ current (I_K) in B cells which is different from those changed by training, but fails to reduce I_A and I_{K-Ca} . Phosphorylase b kinase (an exogenous calcium/calmodulin-dependent kinase) reduces I_A ⁶, but whether I_{K-Ca} and I_{Ca} are changed in the manner of associative training is not yet known. Another protein kinase present in high concentrations in both mammalian brain and molluscan nervous systems is protein kinase C^{7,8}, which is both calcium- and phospholipid-sensitive^{9,10}. We now present evidence that activation of protein kinase C by the tumour promoter phorbol ester (PDB) and intracellular injection of the enzyme induce conductance changes similar to those caused by associative training in *Hermissenda* B cells (that is a reduction of I_A and I_{K-Ca} , and enhancement of I_{Ca}). These results represent the first direct demonstration that protein kinase C affects membrane K⁺ ion conductance mechanisms.

Hermissenda's phototactic behaviour is suppressed when light stimuli are presented in association with rotation (an unconditioned noxious stimulus)^{11,12}. Long-lasting changes in the neural excitability of the B photoreceptors underlie this associative learning^{13,14}. The persistent changes in B photoreceptor excitability include an enhancement of the cells' steady-state depolarizing light response and a decrease in resting membrane conductance^{14,15}. These, in turn, arise from training-produced reductions in a transient voltage-dependent K⁺ current (I_A ²), a delayed calcium-activated K⁺ current (I_{K-Ca} ^{3,4,16}) and enhancement of a slow voltage-dependent calcium current (I_{Ca} ^{3,4}). Pre-

viously, it was suggested that the long-term K⁺ current changes were due to a calcium/calmodulin-mediated phosphorylation process⁶, stimulated by pairing-specific increases in intracellular calcium levels arising from enhanced calcium influx during the acquisition phase of associative learning¹⁷. No mechanism for training-produced changes in I_{Ca} in B cells has been previously proposed or demonstrated.

In B cells, exogenous application of serotonin (10⁻⁵ M), an endogenous neurotransmitter present within the optic ganglia near photoreceptor fine processes and released during associative training¹⁸, produces the same changes of conductance and of light response as associative training¹⁹.

The membrane-soluble tumour promoter 4 β -phorbol 12 β ,13 α -dibutyrate ester (PDB), an activator of protein kinase C²⁰, consistently enhanced the steady-state depolarizing response of type B photoreceptors produced by light when present in concentrations of 1–10 nM (Fig. 1) and decreased the resting membrane conductance (Table 1). These effects were observed within 1–2 min in every cell examined ($n=23$) and persisted for the duration of recording (~1 h) even when followed by extensive washing. Inactive forms of the ester (4 α -phorbol) as well as solvent application (10⁻⁶% acetone, v/v) had no effect. Pretreatment of B cells with 10⁻⁵ M serotonin enhanced the steady-state light response and decreased the resting membrane conductance, as reported previously¹⁹, and diminished the effect of PDB on B cells (Table 1).

Ionophoretic injections of protein kinase C (PKC) mimicked the effects of PDB, serotonin and associative training. PKC enhanced the magnitude of the light response and decreased net membrane conductance in all 13 B cells tested (Fig. 1, Table 1). As was the case for PDB, these effects were extremely persistent, lasting for the duration of recording (~1 h). Ionophoretic injections of heat-inactivated PKC ($n=10$), buffer ($n=5$), or material from a non-PKC peak eluting from the DEAE-cellulose column ($n=3$) were all without effect.

Previous research has determined that the light response of the type B photoreceptor arises from the combined effects of light-, voltage- and calcium-activated ionic currents. Two distinct light-gated currents have been identified: (1) a transient, inward Na⁺ current²¹ ($I_{Na-light}$), which is tetrodotoxin-insensitive, voltage-independent, and underlies the rapid depolarizing limb of the light-induced generator potential and (2) a slower inward current which reflects both a cation-nonspecific conductance increase (J.F., unpublished observations) as well as a light-mediated reduction in a resting K⁺ conductance²². Three voltage-dependent currents have been described: (1) I_A ²³, a transient, rapidly inactivating, 4-aminopyridine (4-AP)-sensitive K⁺ current which is reduced by both associative training² and intracellular calcium load²⁴. In contrast to 'A' currents seen in other gastropod neurones^{25,26}, I_A in *Hermissenda* B cells is activated at membrane potentials near the resting potential of the cell (–50 to –60 mV) and is half-inactivated at a steady-state membrane potential of ~–45 mV²³. In contrast, half-maximal inactivation in both *dorid*²⁵ and *Tritonia*²⁶ neurones occurs at a considerably more negative membrane potential (–70 to –75 mV). (2) I_K , a delayed, sustained, tetraethylammonium (TEA)-sensitive outward K⁺ current observed in many neurone somata²⁷, and (3) I_{Ca} , a small voltage-sensitive inward calcium current²⁸. Also present in B cells is a calcium-activated K⁺ current (I_{K-Ca})^{3,4,28} which resists block by extracellular TEA and 4-AP, contributes to the steady-state depolarizing light response, and is reduced by associative training^{3,4}. We have examined each of these currents to determine which are affected by PDB and PKC.

Under voltage-clamp control, nanomolar concentrations of PDB reduced the fast, rapidly inactivating K⁺ current (I_A) in each of 12 B cells (Fig. 2, Table 1). In eight of these cells PDB application also resulted in a hyperpolarizing shift in the voltage-dependence of inactivation of ~6–8 mV and a more modest shift of 3–4 mV in the voltage-dependence of activation. Because

Fig. 1 PDB and PKC enhance the steady-state depolarizing light response of type B photoreceptors. Type B cells were prepared for recording as in previous studies²⁻⁶ and were ligated so as to remove action- and synaptic-potential-generating areas of the cell. Intracellular recordings were made with a single glass microelectrode (10–30 M Ω , 3 M KCl) connected to a bridge circuit for current injection. Light-evoked depolarizing generator potentials were elicited by 30-s presentations of a 4.0×10^{-4} W cm $^{-2}$ (at 510 nm) light step every 2 min. **a**, Light-induced B-cell response before PDB, in standard artificial sea water (ASW). **b**, Effect of PDB (2 min after bath application). PDB (Sigma) was diluted with acetone to make a 5 mg ml $^{-1}$ stock which was stored frozen at -20°C . The stock was diluted with ASW on the day of the experiment. Final concentrations of PDB and acetone in the bath were $1-10 \times 10^{-9}$ M and $10^{-6}\%$ (v/v), respectively. Application of solvent alone, or inactive forms of the phorbol ester (4 α -phorbol) had no effect. PDB enhanced the steady-state component of the light response. **c**, Light-induced B-cell response in ASW, prior to PKC injection. **d**, Effect of PKC (2 min post-injection). A single 30-s injection of PKC enhanced the steady-state light response of the B cells; a modest reduction of the initial transient depolarizing limb of the generator potential was also often observed. **e**, Summary of data obtained from 23 cells exposed to PDB, 5 cells exposed to acetone solvent (Ac), and 11 cells exposed to 4 α -phorbol (4 α) (C = control). Measurements of the steady-state light-response amplitude were taken at 25 s after light onset and are averages from the three light presentations prior to, and 2 min after, application. **f**, Summary of data from 13 cells exposed to PKC injections, 10 cells exposed to heat-inactivated PKC injections (PKC-HI), and 8 cells exposed to either buffer alone ($n=5$; not shown) or injections of a non-PKC peak eluting from the DEAE-cellulose column ($n=3$, non-PKC). Only PKC injections produced an enhancement of the B cells' light response. All other treatments had no effect. Protein kinase C was partially purified from rat brain using the procedure of Niedel *et al.*³⁵, by ammonium sulphate precipitation and chromatography on a DEAE-cellulose column. The DEAE-cellulose column was washed with 10 ml of the equilibration buffer (20 mM Tris-HCl, pH 7.5/50 mM 2-mercaptoethanol), then a 60-ml linear gradient of 0–0.6 M NaCl in equilibration buffer was applied. The protein peak eluting at 0.3–0.4 M NaCl contained protein kinase C activity, as demonstrated by incorporation of ^{32}P from [γ - ^{32}P]ATP into histone type III-S and ^3H -phorbol diester binding. With incubation conditions as described previously³⁶, in the presence of both Ca^{2+} and phospholipids, enzyme activity resulted in transfer of 54 pM of phosphate per min per mg protein. In the absence of either Ca^{2+} or diacylglycerol, no activity could be detected. Thus, the enzyme was injected in a largely inactive form. The electrode tip was filled via capillary action with PKC in a 20 mM Tris-HCl/350 mM NaCl solution (pH = 7.5). Protein kinase C was iontophoresed as an anion using 1–3 consecutive 15–30-s pulses of anodal (1.0–3.0 nA) current, resulting in the ejection of PKC quantities capable of producing 5×10^{-8} units of activity⁹. A small constant (+0.05 nA) cathodal brake current was applied to prevent leakage of PKC before and after each injection.

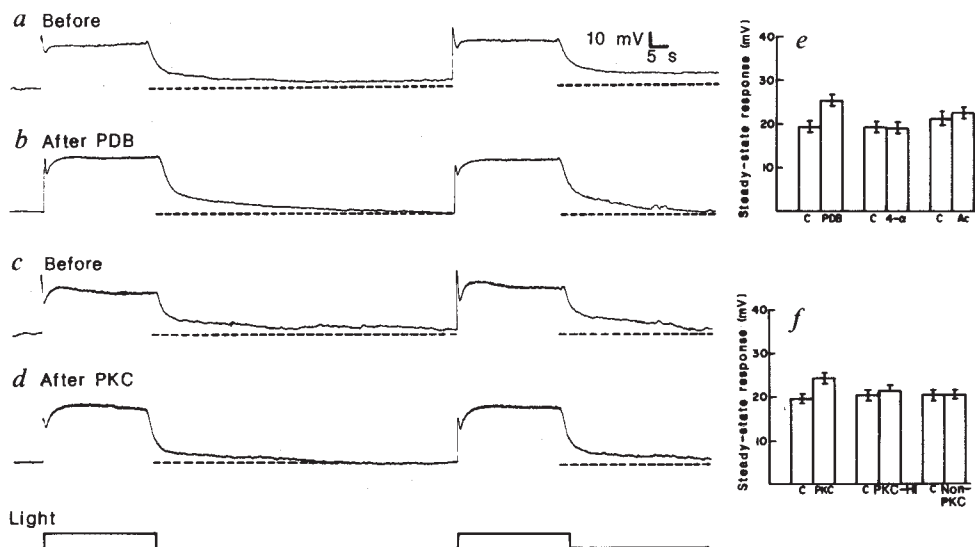
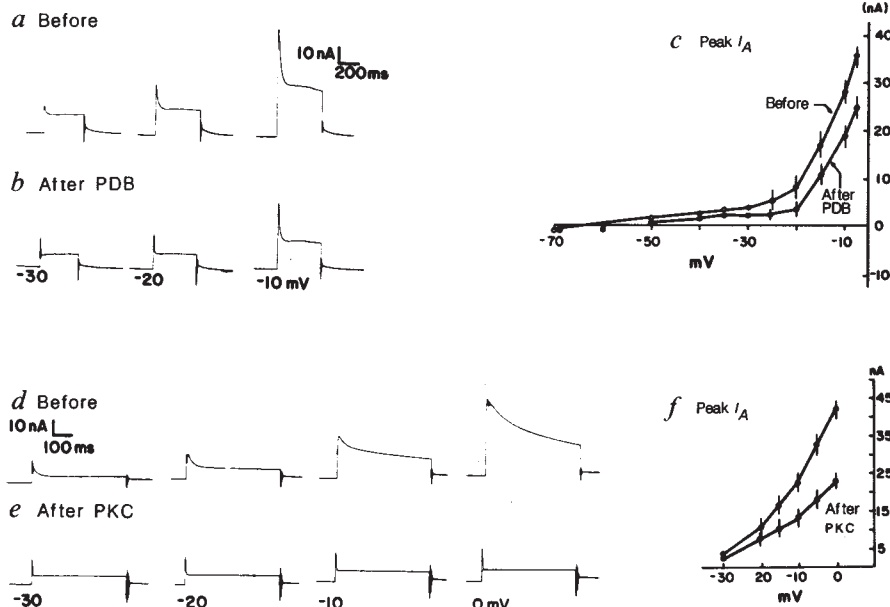


Fig. 2 PDB and PKC reduce I_A . **a**, Outward current records elicited by command steps to potentials more positive than -30 mV in standard ASW exhibit an early (20–50 ms) transient peak (I_A) which then decays to a smaller sustained level of current (comprised of both I_K and I_{K-Ca}). Holding potential (V_h) was -60 mV. Records were obtained before addition of PDB. **b**, Transient peak A current amplitude was reduced by 10^{-9} M bath-applied PDB, as was the steady-state outward current. Records obtained 2 min after PDB application. **c**, Effects of PDB on I_A , averaged for 12 cells. Each data point in the $I-V$ plot represents the average of four or five measurements of transient peak current (absolute amplitude corrected for leakage) at the indicated membrane potentials. Steps were presented at frequencies sufficiently low to preclude any slow-inactivation³⁷ contamination (once per 30 s). **d**, Fast, rapidly inactivating K^+ currents elicited in ASW from a holding potential of -60 mV before PKC injections. **e**, A currents were reduced by a single 30-s (-3.0 -nA anodal current) injection of PKC. Records obtained 5 min after PKC injection. **f**, Average effect of PKC on I_A currents for 10 cells. Error bars represent 1 s.e.m.



Methods. Single synaptically isolated medial and intermediate type B photoreceptors were obtained from the eyes of adult untrained *Hermisenda* by enzymatic treatment with protease (Sigma Type VII) and ligation effected by a razor lesion of the optic nerve. In intracellular microelectrode (3 M KCl, 10–20 M Ω) recording studies (Fig. 1), B cells were characterized by resting potentials > -45 mV, membrane resistance > 25 M Ω , peak and steady-state light-induced (4×10^{-4} W cm $^{-2}$ at 510 nm) generator potentials of > 25 and > 20 mV, respectively, after 6 min of dark adaptation. Two-electrode voltage-clamp methods for the B-cell soma have been extensively described elsewhere²⁻⁶ (3 M KCl electrodes, ~ 6 M Ω for current electrode; 10–20 M Ω for the voltage electrode). Space-clamp control was more than adequate, as judged by the ability to clamp the light-induced generator currents arising on the rhabdomere. In three experiments the cell's isopotentiality was confirmed by inserting a third electrode some distance away from the voltage-recording and current-passing electrode. Membrane potential settling time was 0.2–1.0 ms for all experiments reported. When impaled with two electrodes, cells either satisfied the following criteria or were discarded: (1) resting potential (V_m) > -40 mV; (2) equivalent resting (± 2 mV) and light-induced (peak) generator potentials (> 20 mV) as recorded through current and voltage electrodes; (3) holding currents in ASW ($V_h = -60$ mV) < -2.0 nA. The standard bathing fluid (ASW), unless otherwise noted, was (in mM): 430 Na $^+$, 10 K $^+$, 10 Ca $^{2+}$, 50 Mg $^{2+}$, 10 Tris, 570 Cl $^-$ (pH = 7.6). All experiments were done at room temperature (20°C). All reported current values for Figs 1–3 were corrected for leakage current by extrapolation from a linear portion of the current-voltage relation (10, 20-mV steps from $V_h = -60$ mV).

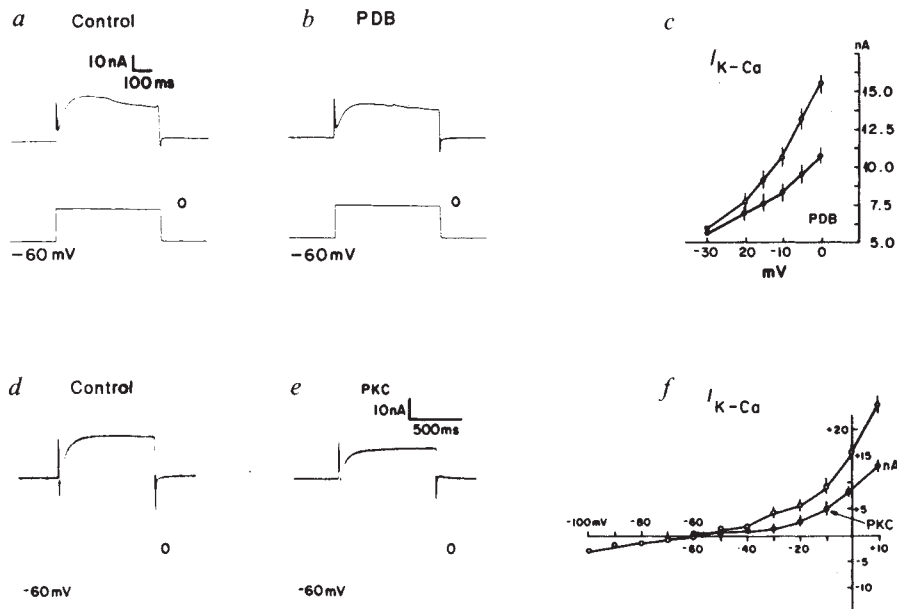


Fig. 3 PDB and PKC reduce I_{K-Ca} . *a*, I_{K-Ca} measured at 0 mV before PDB application. I_{K-Ca} was isolated using 5 mM 4-AP (Sigma) and 100 mM TEA (Aldrich) to block other voltage-dependent K^+ currents²⁸. *b*, PDB reduces I_{K-Ca} . Measurements taken 5 min after bath application of PDB (1×10^{-9} M). Initially, PDB resulted in a slight enhancement of I_{K-Ca} which lasted for 0.5–1.5 min (not shown). *c*, Average PDB effects for 10 cells, 5 min post-PDB (see also Table 1). *d*, I_{K-Ca} measured before PKC injection. *e*, PKC injection reduced I_{K-Ca} ; measurements taken 1.5 min after the PKC injection. *f*, Average results for 8 cells (see also Table 1), 1.5 min post-PKC injection. Error bars, 1 s.e.m.

reduction of I_A by PDB was near-maximal (>90%) in the remaining four preparations, and could not be overcome by steps to more positive potentials or by holding at more negative potentials, it is unlikely that the changes in voltage-dependence account completely for the reduction in the macroscopic current but rather appear to be an additional effect. Bath application of PDB also resulted in a biphasic modulation of I_{K-Ca} . The initial effect (1–3 min) was to enhance I_{K-Ca} ; this was followed, however, by persistent reductions of 20–63% in the current (Fig. 3, Table 1).

The sole voltage-dependent inward current (I_{Ca}) present in B cells is carried by calcium ions²⁸. I_{Ca} is increased by serotonin¹⁹

as well as by associative training^{3,4}. PDB application increased I_{Ca} in the nine cells examined (Fig. 4, Table 1), without obvious effect on the kinetics or voltage-dependence of activation or inactivation. Similar effects on I_{Ca} of *Aplysia* bag-cell neurones by a phorbol ester (tetradecanoyl phorbol 13-acetate) and protein kinase C were described recently by DeRiemer *et al.*²⁹.

Under voltage-clamp control, iontophoretic injections of PKC reduced I_A (Fig. 2, Table 1) and I_{K-Ca} (Fig. 3, Table 1), and enhanced I_{Ca} (Fig. 4, Table 1). Pretreatment of B cells with serotonin reduced these effects of PKC (Table 1). At bath concentration of 1–10 nM, imipramine, which has been reported to inhibit PKC activation³⁰, blocked the enhancement of the B

Table 1 Effects of serotonin, phorbol ester (PDB) and protein kinase C (PKC) on *Hermisenda* type B photoreceptor light response, resting membrane conductance and ionic currents

Agent tested	Steady-state light response (mV)		Net input resistance (M Ω)		I_A (nA) [†]		I_{K-Ca} (nA) [†]		I_{Ca} (nA) [‡]	
	Before	After	Before	After	Before	After	Before	After	Before	After
Serotonin (10^{-5} M)	18.7 ± 1.2	23.7* ± 0.8 (n = 9)	45.3 ± 3.4	71.4* ± 2.9 (n = 9)	31.2 ± 1.4	16.3* ± 1.2 (n = 9)	12.3 ± 0.6	2.1* ± 0.9 (n = 10)	3.5 ± 0.4	5.9* ± 0.4 (n = 8)
PDB (10^{-9} M)	19.1 ± 1.3	25.0* ± 1.3 (n = 23)	42.4 ± 2.8	67.9* ± 2.7 (n = 23)	33.3 ± 2.0	15.7* ± 2.1 (n = 12)	17.2 ± 2.4	9.4* ± 1.8 (n = 10)	3.1 ± 0.5	6.3* ± 0.9 (n = 9)
PDB(serotonin-pretreated) (10^{-9} M, 10^{-5} M)	24.2 ± 1.6	23.9 ± 1.4 (n = 9)	63.8 ± 5.3	66.1 ± 5.2 (n = 9)	20.7 ± 2.5	18.7 ± 2.3 (n = 12)	3.7 ± 1.9	2.9 ± 2.1 (n = 9)	2.5 ± 0.6	3.1 ± 0.7 (n = 8)
4 α -phorbol (10^{-9} M)	19.3 ± 0.9	19.2 ± 1.3 (n = 5)	41.9 ± 2.7	39.2 ± 2.3 (n = 5)	35.3 ± 2.2	32.9 ± 1.4 (n = 4)	15.7 ± 1.3	14.3 ± 1.9 (n = 4)	2.9 ± 1.0	2.7 ± 0.9 (n = 4)
Acetone (10^{-6} %)	21.1 ± 1.7	21.9 ± 1.8 (n = 11)	40.3 ± 2.2	38.7 ± 2.3 (n = 11)	30.0 ± 1.8	28.5 ± 1.5 (n = 4)	18.0 ± 1.2	16.7 ± 1.9 (n = 4)	3.1 ± 0.7	2.9 ± 0.8 (n = 3)
PKC	19.6 ± 0.9	23.5* ± 0.7 (n = 13)	39.3 ± 1.9	57.6* ± 2.1 (n = 13)	31.7 ± 1.8	10.2* ± 0.8 (n = 10)	16.3 ± 1.2	8.7* ± 0.9 (n = 8)	2.9 ± 1.2	5.8* ± 0.9 (n = 7)
PKC(serotonin-pretreated)	23.9 ± 1.3	24.2 ± 1.5 (n = 7)	52.3 ± 4.2	50.9 ± 3.7 (n = 7)	13.0 ± 0.9	12.8 ± 0.7 (n = 5)	4.2 ± 1.2	3.9 ± 0.9 (n = 4)	5.1 ± 0.5	4.9 ± 0.6 (n = 4)
PKC, heat-inactivated	20.1 ± 0.8	21.1 ± 1.2 (n = 10)	37.4 ± 1.2	39.1 ± 2.0 (n = 8)	29.7 ± 1.7	26.4 ± 1.2 (n = 5)	14.7 ± 1.0	11.1 ± 1.2 (n = 4)	3.4 ± 1.5	3.2 ± 2.4 (n = 3)

Values are means $\pm$ s.e.m. All current values are leak-corrected.

* $P < 0.05$, two-tailed t -test; before versus after.

[†] Peak outward current measured at -10 mV, from $V_h = -60$ mV, 5-min post-treatment.

[‡] Steady-state inward current, measured at 0 mV, 5 min post-treatment.

cell's light response and the decrease in conductance caused by PDB and PKC.

These data indicate that activation of protein kinase C in *Hermisenda* B cells can lead to reductions in both voltage- and calcium-activated potassium currents, and enhancement of a voltage-dependent calcium current. The net effect of these conductance changes is an enhancement of the B cell's depolarizing light response and a decrease in resting membrane conductance. Because these effects of PKC are essentially identical to those

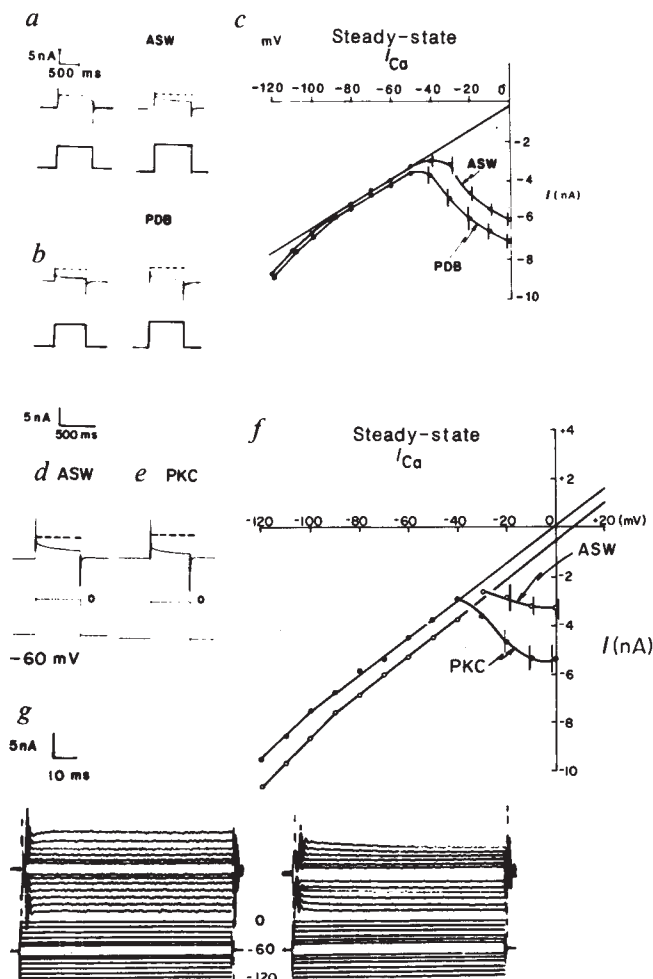


Fig. 4 PDB and PKC enhance I_{Ca} . **a**, I_{Ca} measured at -10 and 0 mV, respectively, before PDB application. I_{Ca} was isolated using 5 mM 4-AP, 100 mM TEA, 10 mM Ba^{2+} for Ca^{2+} , and 300 mM K^+ . In these conditions²⁸, $E_K = \pm 2$ mV. Thus, measurements of the inward current at 0 mV are uncontaminated by any residual K^+ conductance²⁸. Extrapolated leak-current values are denoted by the dotted line. **b**, I_{Ca} measured 2 min after application of 10^{-9} M PDB. I_{Ca} was enhanced. PDB did not result in any change in holding current, however, indicating that it largely affected the voltage-dependent conductance. Leak currents were unaffected. **c**, Average results for nine cells (see also Table 1). Values are for total membrane current. **d**, I_{Ca} measured at 0 mV prior to PKC injection. **e**, PKC injection enhanced I_{Ca} (record obtained 2 min after injection). **f**, Average results for 7 cells (see also Table 1). Values are for total membrane current. PKC injections enhanced I_{Ca} at all potentials more positive than -30 mV. Error bars, 1 s.e.m. **g**, Enhanced inward current is not sodium-dependent. In previous work²⁸, the possibility that Na^+ influx might contribute to the inward current was not directly addressed. Consequently, in further experiments sodium chloride was replaced in Na-free saline by substituting tetramethylammonium chloride (TMA, $n = 6$) or trizma base ($n = 3$) in equimolar concentrations. These substitutions had no effect on the magnitude or kinetics of the inward current, measured in either the presence or absence of PKC injections, except that with prolonged exposure to 0 Na saline, the cells became rather leaky. Shown here is the inward current measured in TMA-ASW before (left) and after (right) PKC injection.

produced by associative training of the animal, which in turn are mimicked by serotonin application, it is attractive to suppose that serotonin-induced PKC activation is one mechanism by which long-term excitability changes in B cells arise from associative training. It is a mechanism which may be distinct from the previously proposed calcium/calmodulin-dependent phosphorylation pathway. Although PKC does require calcium for full activity, data from cell-free enzymatic reaction systems indicate this requirement to be a biochemical rather than a physiological one⁸. In blood platelets 10^{-7} M free calcium has a maximal and saturating effect on PKC activity⁹. Free calcium levels in nerve and neurones are typically already 10^{-7} M³¹, which would be expected to confer a relative independence of PKC from physiological changes in intracellular calcium levels.

Although molecularly distinct, calcium-calmodulin-dependent and PKC-dependent phosphorylation may act in a synergistic fashion to effect long-term current changes in B cells. I_A ²⁴ and I_{K-Ca} ²² in *Hermisenda* B cells are reduced by large increases in the intracellular calcium concentration. Illumination of B cells leads to the release of calcium from intracellular stores in B cells^{17,22}, and may thereby transiently reduce I_A and I_{K-Ca} . In other tissues³², serotonin can lead to mobilization of intracellular calcium, mediated by inositol trisphosphate³³, in addition to its activation of PKC. Thus, it is possible that light and serotonin may act in a synergistic manner in their activation of both protein kinase C and calcium-calmodulin-dependent protein kinase in the B cell. In future studies it will be important to determine whether the effects of PDB and PKC on the ionic conductances are direct, representing PKC-dependent phosphorylation of ionic channels, or indirect, perhaps through an interaction with calcium/calmodulin or cyclic nucleotide pathways (as has been demonstrated in oocytes³⁴).

This research was supported by NSF grant BNS-8316707 (J.F.).

Received 3 June; accepted 15 November 1985.

- Nestler, E. J. & Greengard, P. *Nature* **305**, 583-588 (1983).
- Alkon, D., Lederhendler, I. & Shoukimas, J. *Science* **215**, 693-695 (1982).
- Farley, J. & Alkon, D. L. *Soc. Neurosci. Abstr.* **9**, 167 (1983).
- Farley, J., Sakakibara, M. & Alkon, D. L. *Soc. Neurosci. Abstr.* **10**, 270 (1984).
- Alkon, D. L., Acosta-Urquidí, J., Olds, J., Kuzma, G. & Neary, J. T. *Science* **219**, 303-306 (1983).
- Acosta-Urquidí, J., Alkon, D. L. & Neary, J. T. *Science* **224**, 1254-1257 (1984).
- Minakuchi, R., Takai, Y., Yu, B. & Nishizuka, Y. *J. Biochem., Tokyo* **89**, 1651-1654 (1981).
- Nishizuka, Y. *Science* **225**, 1365-1370 (1984).
- Kishimoto, A., Takai, Y., Mori, T., Kikkawa, W. & Nishizuka, Y. *J. Biol. Chem.* **255**, 2273-2276 (1980).
- Berridge, M. J. *Molec. cell. Endocr.* **24**, 115-140 (1981).
- Crow, T. & Alkon, D. L. *Science* **201**, 1239-1241 (1978).
- Farley, J. & Alkon, D. L. *Science* **209**, 412-414 (1980).
- Farley, J., Richards, W., Ling, L., Liman, E. & Alkon, D. L. *Science* **221**, 1201-1203 (1983).
- Farley, J. & Alkon, D. L. *J. Neurophys.* **48**, 785-807 (1982).
- West, A., Barnes, E. & Alkon, D. L. *J. Neurophys.* **48**, 1243-1255 (1982).
- Grover, L. & Farley, J. *Soc. Neurosci. Abstr.* **11**, 788 (1985).
- Connor, J. & Alkon, D. L. *J. Neurophys.* **51**, 745-752 (1984).
- Auerbach, S., Grover, L. & Farley, J. *Soc. Neurosci. Abstr.* **11**, 481 (1985).
- Wu, R. & Farley, J. *Soc. Neurosci. Abstr.* **10**, 620 (1984).
- Castagna, M. *et al. J. Biol. Chem.* **257**, 7847-7851 (1979).
- Alkon, D. L. *Science* **205**, 810-816 (1979).
- Alkon, D. L. & Sakakibara, M. *Soc. Neurosci. Abstr.* **10**, 10 (1984).
- Shoukimas, J. J. & Alkon, D. L. *Soc. Neurosci. Abstr.* **6**, 17 (1980).
- Alkon, D. L., Shoukimas, J. J. & Heldman, E. *Biophys. J.* **40**, 245-250 (1982).
- Connor, J. A. & Stevens, C. F. *J. Physiol., Lond.* **213**, 21-30 (1971).
- Thompson, S. J. *J. Physiol., Lond.* **265**, 465-488 (1977).
- Adams, D. J., Smith, S. J. & Thompson, S. H. A. *Rev. Neurosci.* **3**, 141-167 (1980).
- Alkon, D. L., Farley, J., Sakakibara, M. & Hay, B. *Biophys. J.* **46**, 605-614 (1984).
- DeRiemer, S. A., Strong, J. A., Albert, K. A., Greengard, P. & Kaczmarek, L. K. *Nature* **313**, 313-316 (1985).
- Mori, T., Takai, Y., Minakuchi, R., Yu, B. & Nishizuka, Y. *J. Biol. Chem.* **255**, 8378-8380 (1980).
- Baker, P. F. *Prog. biophys. molec. Biol.* **24**, 177-223 (1972).
- Berridge, M. J. *Biochem. J.* **212**, 849-858 (1983).
- Berridge, M. J. & Irvine, R. F. *Nature* **312**, 315-321 (1984).
- Dascal, N., Lotan, I., Gillo, B., Lester, H. A. & Lass, Y. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6001-6005 (1985).
- Niedel, J. E., Kuhn, L. J. & Vandenbark, G. R. *Proc. natn. Acad. Sci. U.S.A.* **80**, 36-40 (1983).
- Wise, B. C., Raynor, R. L. & Kuo, J. F. *J. Biol. Chem.* **257**, 8481-8488 (1982).
- Aldrich, R. W. Jr, Getting, P. A. & Thompson, S. H. *J. Physiol., Lond.* **291**, 531-544 (1979).

Differential expression of α -fetoprotein genes on the inactive X chromosome in extraembryonic and somatic tissues of a transgenic mouse line

R. Krumlauf*, V. M. Chapman†, R. E. Hammer‡, R. Brinster‡ & S. M. Tilghman*

* Institute for Cancer Research, 7701 Burholme Ave, Philadelphia, Pennsylvania 19111, USA

† Roswell Park Memorial Institute, 666 Elm St, Buffalo, New York 14263, USA

‡ Laboratory of Reproductive Physiology, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

During development of the female mouse embryo, one of the two X chromosomes is inactivated in a random manner in most cell lineages¹. However, in the extraembryonic trophectoderm and primary endoderm lineages there is preferential inactivation of the paternally derived X chromosome²⁻⁴. The inactivated X chromosomes of the extraembryonic and somatic tissues appear equally inactive at the level of the expression of X-linked genes²⁻⁷. However, there are differences in the timing of their replication⁸⁻¹⁰ and the extent of DNA modification as determined by gene transfer¹¹⁻¹⁵. The identification of transgenic animals carrying X-linked modified α -fetoprotein (AFP) genes allowed us to examine whether the inactivation process extends to an autosomal gene which is normally expressed at high levels in specific extraembryonic and somatic cells, and if so, whether the inactivation process is different in these two tissues. Our results demonstrate that the X-linked AFP genes were expressed on the inactive X chromosome in the visceral endoderm of the yolk sac but not in fetal liver. Thus, the transcriptional activity of the AFP minigene on the inactive X chromosome is dependent on the tissue in which it resides, and most probably reflects differences in the nature of the maintenance of the inactive state of the extraembryonic and embryonic X chromosomes.

We have shown previously that transgenic mice carrying exogenous AFP genes express these genes in a tissue-specific and temporally appropriate manner in the visceral endoderm of the yolk sac, the fetal liver and the fetal gastrointestinal tract¹⁶. During breeding experiments with one line derived from a male F₁ of the founder mouse, 47-3, we observed that the two to three tandem copies of the AFP minigene were segregating with the X chromosome. In analysing more than 1,200 progeny derived from five F₁ transgenic males, all females and no males carried the AFP minigenes. In contrast, females were able to transmit the genes equally to females and males. A partial pedigree is shown in Fig. 1.

To determine whether the AFP minigenes on the paternally derived X chromosome were preferentially inactivated in the visceral endoderm of the yolk sac, a tissue in which AFP messenger RNA constitutes 20% of the mRNA pool, animals were obtained by caesarean section at 14-16 days of fetal development from matings between (C57BL/6 $\times$ SJL)F₁ mice with either male or female transgenic mice. Transgenic fetuses were identified by hybridizing fetal liver genomic DNA to a first exon-specific AFP probe able to detect both the endogenous and the introduced AFP genes^{16,17}. The sex of the fetuses was determined by hybridization of genomic DNA to a retroviral probe (M270) that can detect sequences on the Y chromosome¹⁸ (data not shown).

Total yolk sac poly(A)⁺ RNA from female transgenic progeny was isolated, fractionated by electrophoresis and hybridized to an AFP probe. The introduced AFP minigene is an internally deleted AFP gene^{16,17} that generates a correctly initiated and processed 600-nucleotide poly(A)⁺ RNA readily distinguished

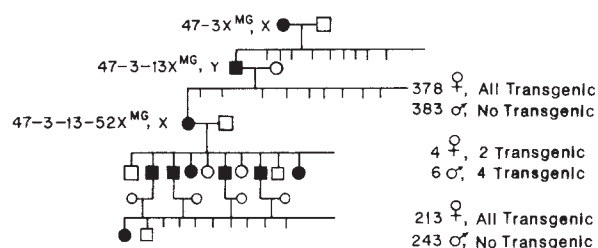


Fig. 1 Simplified pedigree of the 47-3-13 line showing segregation of the X-linked AFP minigenes. The founder mouse (47-3) was mated to (C57BL/6 $\times$ SJL)F₁ males and one of her male progeny (47-3-13) was mated to BALB/cJ mice. The sex of all subsequent progeny was determined by external genitalia and/or tail DNA analysis¹⁶ using the retroviral probe (M720) which is able to detect sequences on the Y chromosome¹⁸. Males are indicated by squares and females by circles. Transgenic animals (closed symbols) were identified with an AFP hybridization probe in the manner shown in Fig. 4b. MG refers to the AFP minigene.

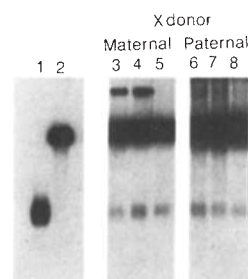


Fig. 2 Expression in yolk sac of the AFP minigenes located on the inactive paternal or active maternal X chromosome. Lanes 1 and 2 are control RNAs denoting the migration of AFP minigene (lane 1) and authentic AFP mRNA (lane 2). Transgenic females or males were mated to (C57BL/6 $\times$ SJL)F₁ animals, and transgenic female progeny were identified on days 14-16 post-fertilization by hybridization of fetal liver and yolk sac DNAs, isolated according to published procedures¹⁶, to AFP- and Y-chromosome¹⁸-specific sequences. Poly(A)⁺ RNA was isolated from individual yolk sacs of progeny derived from transgenic females (lanes 3-5) and males (lanes 6-8), fractionated on denaturing formaldehyde 1.7% agarose gels and transferred to nitrocellulose. The RNAs were hybridized to an AFP-specific single-stranded RNA probe as detailed previously¹⁶.

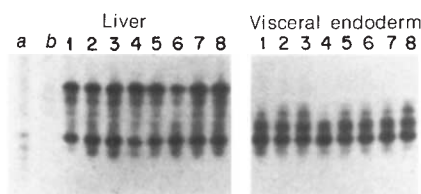


Fig. 3 Differential expression of HPRT isozymes in the yolk sac and liver of transgenic females. The pattern of expression of HPRT isozymes in eight transgenic female livers (lanes 1-8, left) and visceral endoderms (right) was determined. Lanes a and b are control samples for the HPRT-A and HPRT-B isozymes, respectively. Male transgenic animals carrying the *pgk-1*^b and *hprt*^b alleles were mated to C3H/He female congenic mice heterozygous for the *pgk-1*^{a/b} and *hprt*^{a/b} alleles. Fetuses were obtained by caesarian section at 14-16 days' gestation and the livers and yolk sacs removed. The visceral endoderm of the yolk sac was separated from the mesodermal layer by incubating the tissue in an enzyme mixture (2.5% pancreatin, 0.5% trypsin, 1 $\times$ pucks saline A) at 4 $^{\circ}$ C for 1 h as described previously^{5-7,9,15,28}. The tissue was transferred to phosphate-buffered saline at 4 $^{\circ}$ C and the layers were separated with forceps under a dissecting microscope. In some cases DNA and RNA were isolated from the tissues and analysed, and in others HPRT and PGK isozyme analysis was performed as described previously^{12,15,19,20}. Only progeny which exhibited the informative *hprt*^{a/b} and *pgk-1*^{a/b} genotype (16/97) are shown.

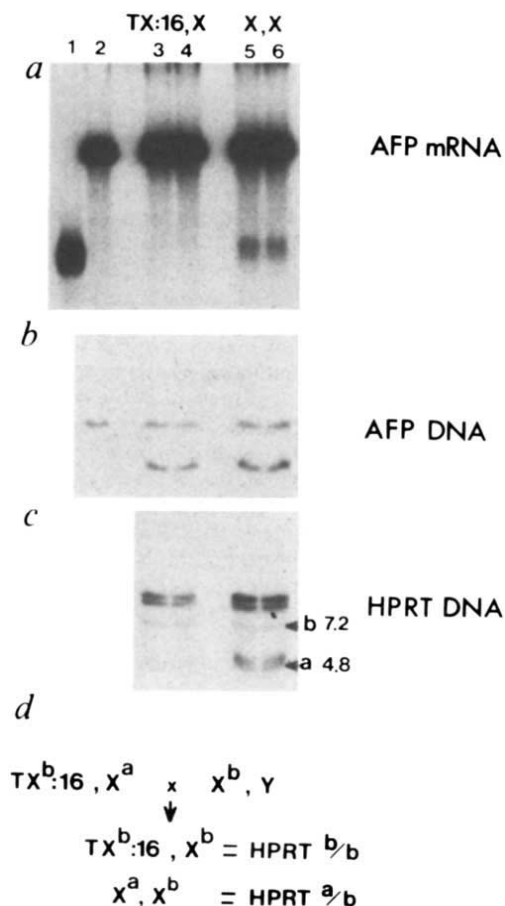


Fig. 4 Expression of the X-linked AFP minigenes in the livers of animals bearing the Searle's translocation. *a*, Poly(A)⁺ RNA was isolated from day 1 neonatal animals and analysed after gel electrophoresis by hybridization to an AFP-specific probe. Lane 1 is a control for the migration of the minigene mRNA and lane 2 is a control RNA isolated from a non-transgenic animal. Lanes 3 and 4 are from two transgenic animals carrying the Searle's translocation and lanes 5 and 6 from two normal XX transgenic littermates. *b*, Hybridization of genomic DNA isolated from livers of the same animals as in *a* to an AFP-specific probe. The faster-migrating band indicates the presence of the X-linked minigene in lanes 3–6. *c*, Hybridization of the same genomic DNA in lanes 3–6 with an *hprt* cDNA probe that distinguishes between the *hprt*^a and *hprt*^b alleles. *d*, A diagrammatic representation of the mating and predicted genotypes of the female offspring.

Methods. Female mice carrying Searle's translocation²¹ have the *pgk-1*^b and *hprt*^b alleles on the translocated portion of the X chromosome, and the *pgk-1*^a and *hprt*^a alleles on the normal X chromosome. These animals were mated to transgenic males carrying the X-linked *hprt*^b and *pgk-1*^b alleles and the AFP minigenes. Progeny were killed on day 1 of neonatal life and RNA and DNA were isolated from the liver as described previously¹⁶. The DNA samples (10 µg) were restricted with *Hind*III, electrophoresed through a 1% agarose gel, transferred to nitrocellulose and hybridized with either a nick-translated AFP probe or an *hprt* cDNA probe as described elsewhere¹⁶. The 4.8-kb *Hind*III fragment is a restriction fragment length polymorphism associated with the *hprt*^a allele used in this cross (F. Berger and V.M.C., unpublished results). It is present on the intact X chromosome of the t(X; 16H) dam of these offspring. Thus *hprt*^{a/b} progeny are X/X while *hprt*^b progeny are probably t(X; 16H) since *hprt* maps within 5 centimorgans of the translocation breakpoint. The poly(A)⁺ RNA samples (~1 µg) were fractionated by electrophoresis through a denaturing formaldehyde agarose gel, transferred to nitrocellulose and hybridized with a single-stranded AFP probe as described previously¹⁶.

from the 2.2-kilobase (kb) endogenous mRNA. Figure 2 shows that the AFP minigenes were expressed in the extraembryonic yolk sac, irrespective of whether they derived from the paternal (inactive) or maternal (active) X chromosome. Partial paternal X inactivation was suggested by a general two- to three-fold higher level of expression from the AFP minigenes when they were present on the maternally derived X chromosome.

To ensure that the paternal X chromosome was actually inactivated, male transgenic animals carrying the X-linked phosphoglycerate kinase-B (*pgk-1*^b) and hypoxanthine phosphoribosyl transferase-B (*hprt*^b) alleles were mated to C3H/He female congenic mice carrying the *pgk-1*^a and *hprt*^a alleles derived from wild mice^{19,20}. Fetuses were collected at 14–16 days of gestation and the visceral endoderm of the yolk sac was isolated by dissection following enzymatic digestion^{5,7,9,15}. Assays to distinguish the PGK and HPRT isoenzyme^{15,19,20} and to quantitate AFP minigene mRNA were performed on both fetal liver and visceral endoderm samples. For eight of the appropriate females, both the A and B isozymes of HPRT were expressed in liver, whereas only the A isozyme from the maternally derived X chromosome was expressed in the visceral endoderm of the same animals (Fig. 3). Similar results were obtained for PGK (data not shown). However, as before, the AFP minigenes were expressed irrespective of their origin. Therefore, the ability of the minigenes to escape X-chromosome inactivation cannot be attributed to their presence interfering with the inactivation of the entire chromosome.

This experiment does not exclude the possibility that the AFP minigenes have integrated into a region of the X chromosome which normally escapes inactivation. On the other hand, the activity of X-linked AFP genes in the extraembryonic lineage could be a unique property of the X chromosome of that tissue. One way to distinguish between these possibilities is to examine the activity of the AFP minigenes on the inactive X chromosome in a somatic tissue, such as fetal liver. To do this we took advantage of a strain of mice carrying one copy of a balanced X:autosome translocation (t(X; 16H)) termed Searle's translocation²¹. In all females that carry the translocated X, the normal X chromosome donated by the male is inactive^{5,7,10,12,15,22}, presumably to prevent the partial monosomy of chromosome 16 that would otherwise occur. Thus, by crossing a transgenic male to a Searle's female, we will generate female progeny in which the X chromosome bearing the AFP minigene is inactive in all tissues. Normal XX females which will also result from this cross display random X inactivation and serve as controls. A diagrammatic representation of the mating and expected genotypes is shown in Fig. 4d.

Progeny were killed on day 1 of neonatal life and DNA and RNA were isolated from the liver, a tissue where AFP mRNA levels are ~10% of total poly(A)⁺ mRNA. Transgenic animals were identified by DNA hybridization to an AFP probe (Fig. 4b) and normal XX females were distinguished from translocation-bearing females by using a restriction fragment length polymorphism that distinguishes between the *hprt*^a and *hprt*^b alleles (Fig. 4c). Analyses of AFP minigene expression in representative transgenic females are shown in Fig. 4a. The AFP minigenes were not expressed in liver on the inactive X of the females bearing the translocation, although expression was observed in normal XX littermates. Therefore, the transcriptional activity of the AFP minigenes on the inactive X chromosome is dependent on the tissue in which the chromosome resides.

The lack of expression of the minigenes on the inactive X chromosome in liver could be anticipated from studies on X chromosome:autosome translocations, where the inactive state extended past the breakpoint into the autosomal genes^{23–25}. The surprising result is the continued expression in the visceral endoderm, which suggests that the inactivation mechanism can discriminate between an X-linked gene and an autosomal gene. This difference may not be a property of the X chromosome, but rather reflects differences in the modes of regulation of the

AFP gene itself in these tissues; for example, the concentration of critical transcription factors may be high enough in the visceral endoderm to gain access to the gene even on the inactive X.

Alternatively and more likely, these findings can be interpreted as support for the notion that there are functional differences between the inactive X chromosome in extraembryonic versus embryonic tissues. It has been argued that these differences reside at the level of the maintenance^{12,15} rather than the establishment of inactivation, as by the criteria of gene activity both chromosomes are equally inactive. The X-linked AFP genes were partially inactivated (two-three-fold) on the paternal X in visceral endoderm, which is consistent with this idea.

One candidate to explain the differences between the two inactive chromosomes is differences in DNA modification, particularly methylation. This argument is based on gene transfer experiments which demonstrated that the activity of the *hprt* gene can be transferred to a recipient cell from the inactive X chromosome of embryonic origin after treatment of the cell with 5-azacytidine, unlike the extraembryonic derived *hprt* gene, which can be transferred without drug treatment. The genomes of the extraembryonic lineages in general are undermethylated relative to other somatic tissues^{26,27}, a condition that extends from structural genes independently of whether they are expressed to satellite DNA. Using the isoschizomers *Msp*I and *Hpa*II, which monitor the methylation status of a subset of CpG sites in the AFP minigene integration site, we have not detected any significant differences between the degrees of methylation of the locus on the inactive X in visceral endoderm and liver (data not shown). These experiments argue against global methylation differences as an explanation for continued expression of the minigenes on the visceral endoderm-derived inactive X chromosome.

In conclusion, the differential expression of the AFP minigenes on the inactive X chromosome in the visceral endoderm and liver suggests that the mechanism of maintenance of the inactive state is discriminating between normally X-linked genes and an autosomal gene in the former tissue but not in the latter.

We thank E. Eicher for providing the M720 probe, T. Caskey for the *hprt* cDNA probe, A. McLaren, M. Monk, B. Hogan and T. Vogt for valuable discussions and Marie Estes for the preparation of the manuscript. R.K. and R.E.H. were supported by NIH postdoctoral training grants CA09035 and HD07155, respectively. This work was also supported by NIH grants HD17321, GM24125, CA06927, CA28050 and RR05539 and an appropriation from the Commonwealth of Pennsylvania.

Received 19 September; accepted 31 October 1985.

1. Lyon, M. F. *Nature* **190**, 372-373 (1961).
2. Takagi, N. & Sasaki, M. *Nature* **256**, 640-642 (1975).
3. West, J. D., Frels, W. I., Chapman, V. M. & Papaioannou, V. E. *Cell* **12**, 873-882 (1977).
4. Papaioannou, V. E. & West, J. D. *Genet. Res. Camb.* **37**, 183-197 (1981).
5. McMahon, A., Forsten, M. & Monk, M. *J. Embryol. exp. Morph.* **64**, 251-258 (1981).
6. McMahon, A., Forsten, M. & Monk, M. *J. Embryol. exp. Morph.* **74**, 207-220 (1983).
7. McMahon, A. & Monk, M. *Genet. Res. Camb.* **41**, 69-83 (1983).
8. Takagi, N. *Expl. Cell Res.* **86**, 127-135 (1974).
9. Sugawara, O., Takagi, N. & Sasaki, M. *Chromosoma* **88**, 133-138 (1983).
10. Distche, C. M., Eicher, E. M. & Latt, S. A. *Expl. Cell Res.* **133**, 357-362 (1981).
11. Liskay, R. M. & Evans, R. J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4895-4898 (1980).
12. Chapman, V. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 5357-5361 (1982).
13. Lester, S. C., Korn, N. J. & DeMars, R. *Somatic Cell Genet.* **8**, 265-284 (1982).
14. Venolia, L. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 2352-2354 (1982).
15. Kratzer, P. G., Chapman, V. M., Lambert, H., Evans, R. E. & Liskay, R. M. *Cell* **33**, 37-42 (1983).
16. Krumlauf, R., Hammer, R. E., Tilghman, S. M. & Brinster, R. L. *Molec. cell. Biol.* **5**, 1639-1648 (1985).
17. Scott, R. W., Vogt, T. F., Croke, M. E. & Tilghman, S. M. *Nature* **310**, 562-567 (1984).
18. Phillips, S. J., Birkenmeier, E. H., Callahan, R. & Eicher, E. M. *Nature* **297**, 241-242 (1982).
19. Nielsen, J. T. & Chapman, V. M. *Genetics* **87**, 319-325 (1977).
20. Chapman, V. M., Kratzer, P. G. & Quarantillo, B. A. *Genetics* **103**, 785-795 (1983).
21. Lyon, M. F., Searle, A. G., Ford, C. E. & Ohno, S. *Cytogenetics* **3**, 306-323 (1964).
22. Takagi, N. *Chromosoma* **81**, 439-445 (1980).
23. Russell, L. B. & Montgomery, C. S. *Genetics* **64**, 281-312 (1970).
24. Russell, L. B. *Science* **140**, 976-978 (1963).
25. Ohno, S. & Lyon, M. F. *Chromosoma* **16**, 90-100 (1965).
26. Young, P. R. & Tilghman, S. M. *Molec. Cell Biol.* **4**, 898-907 (1984).
27. Sanford, J. P., Chapman, V. M. & Rossant, J. *Trends Genet.* **1**, 89-93 (1985).
28. Levak-Svajger, B., Svajger, A. & Skreb, N. *Experientia* **25**, 1311-1312 (1969).

The *neu* oncogene encodes an epidermal growth factor receptor-related protein

Cornelia I. Bargmann, Mien-Chie Hung & Robert A. Weinberg

Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142 and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

The *neu* oncogene is repeatedly activated in neuro- and glioblastomas derived by transplacental mutagenesis of the BDIX strain of rat with ethylnitrosourea¹. Foci induced by the DNAs from such tumours on NIH 3T3 cells contain the *neu* oncogene and an associated phosphoprotein of relative molecular mass 185,000 (p185)². Previous work has shown that the *neu* gene is related to, but distinct from, the gene encoding the EGF receptor (*c-erb-B*)^{3,4}. Here we describe a *neu* complementary DNA clone isolated from a cell line transformed by this oncogene; the clone has biological activity in a focus-forming assay. The nucleotide sequence of this clone predicts a 1,260-amino-acid transmembrane protein product similar in overall structure to the EGF receptor. We found that 50% of the predicted amino acids of *neu* and the EGF receptor are identical; >80% of the amino acids in the tyrosine kinase domain are identical. Our results suggest strongly that the *neu* gene encodes the receptor for an as yet unidentified growth factor.

Transcripts of the *neu* gene were identified in NIH 3T3 cells transformed by *neu*. Polyadenylated RNA was isolated from B104-1-1, an NIH 3T3-derived cell line transformed by DNA from the B104 rat neuroblastoma cell line and known to express a high level of p185 (ref. 2). RNA was also isolated from NIH 3T3 cells transformed by a *ras* oncogene and from Rat-1 cells, a rat fibroblast line which expresses a protein immunologically related to p185 (ref. 3). This protein, which is also of $M_r \sim 185,000$, is probably the product of the normal allele of the *neu* gene. A Northern blot of these RNAs was probed with R4E, a 4-kilobase (kb) *Bam*HI clone isolated from a genomic rat liver library by hybridization with part of the *v-erb-B* oncogene. This clone corresponds to the putative kinase domain of the rat *neu* gene (data not shown).

Under stringent conditions, this probe hybridized with RNAs of 4.4, 5.0 and 9 kb present in the *neu*-transformed NIH 3T3 cell line that were not present in the *ras*-transformed cell line (Fig. 1a,b). The smaller of these RNAs appeared also to be present in Rat-1 cells (Fig. 1c). These data strongly suggest that these messenger RNAs are the transcripts derived from the rat *neu* gene present in B104-1-1 cells and not the endogenous mouse gene. Quantitation of RNA levels by hybridization with a single-stranded RNA probe specific for *neu* indicated that these *neu* transcripts represent between 0.01 and 0.05% of the polyadenylated RNA in B104-1-1 cells.

A cDNA library was prepared from B104-1-1 polyadenylated RNA using standard *S*₁ snapback techniques⁵. Nine recombinant plasmids reactive with a *neu* probe were isolated, eight of which were derived from the same family of mRNAs as determined by restriction mapping. The largest of these, designated *neuc(t)34*, had an insert of 4.8 kb and was used for all subsequent studies.

In Northern analysis this cDNA clone reacted with all three of the mRNAs detected by the genomic clone (data not shown) and with no additional transcripts. A 400-base pair (bp) fragment from the extreme 3' end of *neuc(t)34* was used as a probe on an identical Northern filter and, as shown in Fig. 1d-f, it preferentially detected the 5.0- and 9-kb mRNAs in these cell lines. This suggests that the difference between the 4.4- and the 5.0-kb mRNAs resides at the 3' end of the gene and that the *neuc(t)34* cDNA is derived from the 5.0-kb mRNA.

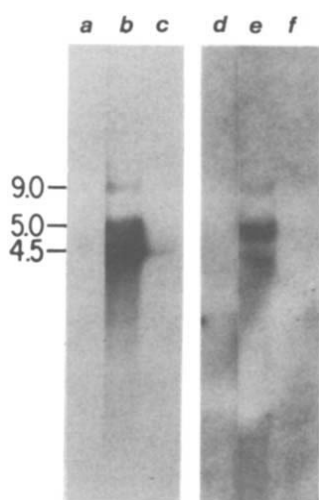


Fig. 1 Northern blot analysis of *neu*-related RNAs. Poly(A)⁺ RNA (10 µg per lane) was hybridized with radiolabelled DNA from a genomic rat *neu* clone (*a-c*) or from the extreme 3' end of the cDNA clone *neuc(t)34* (*d-f*). RNA was extracted from Ki-ras-transformed NIH 3T3 cell line A5-2 (*a, d*), *neu*-transformed NIH 3T3 cell line B104-1-1 (*b, e*), or the Rat-1 fibroblast line (*c, f*) by standard methods⁷, followed by oligo(dT) selection and separated on a formaldehyde-containing 1% agarose gel. *Hind*III-cleaved λ DNA was run as a marker. The RNA was transferred to nitrocellulose and hybridized in 50% formamide, 5×SSCPE, 1×Denhardt's, 100 µg ml⁻¹ boiled DNA and 10% dextran sulphate containing 5×10⁶ c.p.m. of either a 500-bp *Bgl*II-*Stu*I fragment of genomic clone R4E (described in the text) or a 1.5-kb *Bam*HI fragment of *neuc(t)34* nick-translated to a specific activity of 10⁹ c.p.m. µg⁻¹; the nitrocellulose filters were washed at 65°C in 2×SSCPE, 0.1% SDS and exposed to Kodak X-omat R film for 3 days at -70°C using an intensifying screen.

The 4.8-kb insert of *neuc(t)34* was introduced into the expression vector pSV2 (ref. 6) and tested for biological activity by transfection into NIH 3T3 cells. This clone was highly active in causing focus formation, yielding ~1,000 foci per µg of transfected DNA. Cell lines isolated from such foci contain a protein that is immunologically related to and co-migrates with the authentic p185 of B104-1-1 cells (data not shown). We conclude that *neuc(t)34* is a biologically active clone of the *neu* gene present in B104-1-1 cells and encodes a full-length (or nearly full-length) p185 gene product.

Both strands of the coding region of *neuc(t)34* were sequenced. Most parts of the gene were sequenced at least once each by the dideoxy and Maxam-Gilbert methods⁷⁻⁹. Figure 2 shows the DNA sequence of the *neu* cDNA and the predicted amino-acid sequence of its protein product; this predicted protein sequence is compared with that of the human EGF receptor in Fig. 3.

Neuc(t)34 has an open reading frame of 3,780 nucleotides and encodes a putative primary translation product of 1,260 amino acids that has a total *M_r* of 139,221, significantly smaller than the observed size of the *neu* gene product (*M_r* = 185,000). Glycosylation accounts for 15,000 daltons of the discrepancy (D. Stern and R.A.W., in preparation), and other post-translational modifications have not been excluded.

Sixteen of the 19 amino acids following the first ATG in the sequence are hydrophobic residues, consistent with the formation of a signal sequence for insertion into the membrane¹⁰. Residues 658-680 of the sequence are also highly hydrophobic, which suggests that this region represents a transmembrane domain. The location of these sequences and their similarity to sequences in the evolutionarily related EGF receptor¹¹ lead us to propose that p185 is organized similarly to the EGF receptor. Thus, the first few amino-acid residues of p185 would represent a cleaved signal sequence, the next 640 residues would include

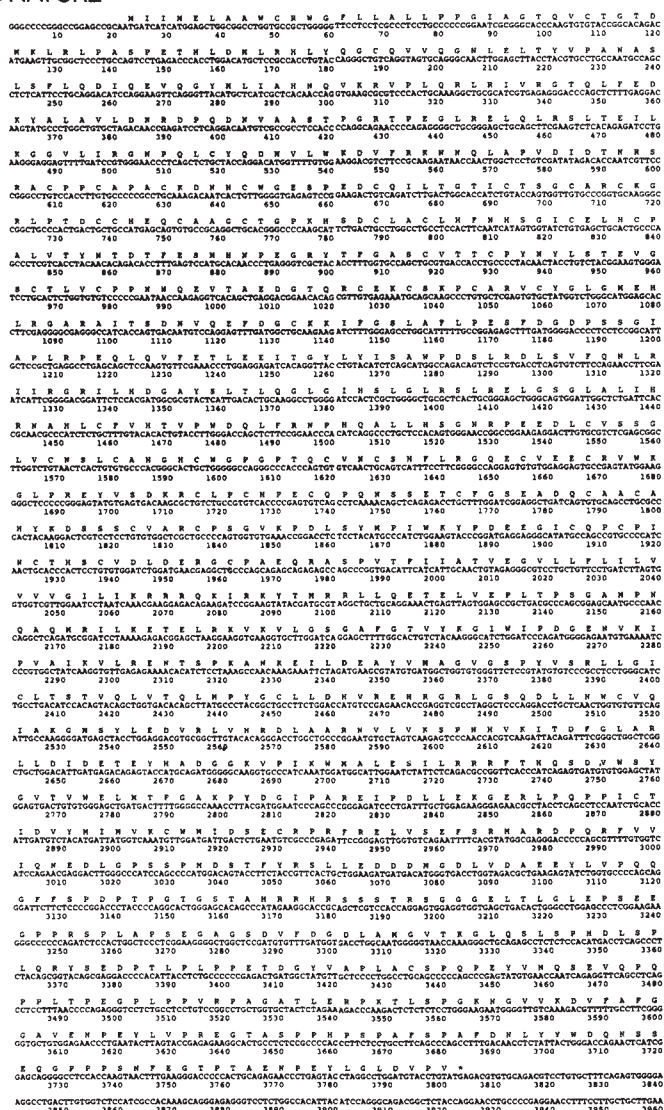


Fig. 2 DNA sequence of *neu*. Maxam-Gilbert sequencing⁹ and M13 chain-termination sequencing^{10,11} have been described elsewhere. Nucleotides 1-5 were generated by the tailing of the cDNA and do not represent sequences present in the RNA.

Methods. Poly(A)⁺ RNA (5 µg) isolated as described in Fig. 1 legend was denatured in 12.5 mM methyl mercury hydroxide and reverse-transcribed in 50 mM Tris-HCl pH 8.3, 100 mM KCl, 10 mM MgCl₂, 5 mM dithiothreitol (DTT) and 0.5 mM of each dNTP with 50 µg ml⁻¹ oligo(dT) (Collaborative Research), 120 U RNase (Promega Biotech) and 150 U reverse transcriptase (Boehringer-Mannheim) for 90 min at 42°C. Nucleic acids were precipitated and the second-strand polymerization carried out in 50 mM KCl, 10 mM MgCl₂, 50 mM Tris pH 7.5, 1 mM DTT and 0.5 mM of each dNTP with 50 U DNA polymerase I large fragment (New England Biolabs) for 12 h at 16°C after a second methyl mercury hydroxide denaturation. Hairpin loops were nicked with S₁ nuclease (600 U; Sigma) in 50 mM NaOAc pH 4.5, 0.2 M NaCl and 2 mM ZnSO₄ for 20 min at 37°C. Double-stranded cDNA was size-selected over a BioGel A15 column (BioRad) and tailed using 30 U terminal transferase (BRL) in 200 mM potassium cacodylate, 2 mM DTT, 0.5 mM CoCl₂, 20 mM Tris pH 6.9, 600 µg ml⁻¹ bovine serum albumin, and 0.1 mM dCTP for 2 min at 30°C, phenol-extracted, ethanol-precipitated and annealed into dG-tailed *Pst*-cut pBR322 (BRL). 50,000 independent colonies were derived after transformation of cDNA into competent *Escherichia coli* and plating onto tetracycline-containing plates; these colonies were screened with a nick-translated fragment of a genomic clone of rat *neu* consisting of a 500-bp *Bgl*II to *Stu*I fragment which encompasses two exons of the *neu* tyrosine kinase domain. Nine recombinant plasmids reactive with this probe were purified; *neuc(t)34*, the largest, was sequenced.

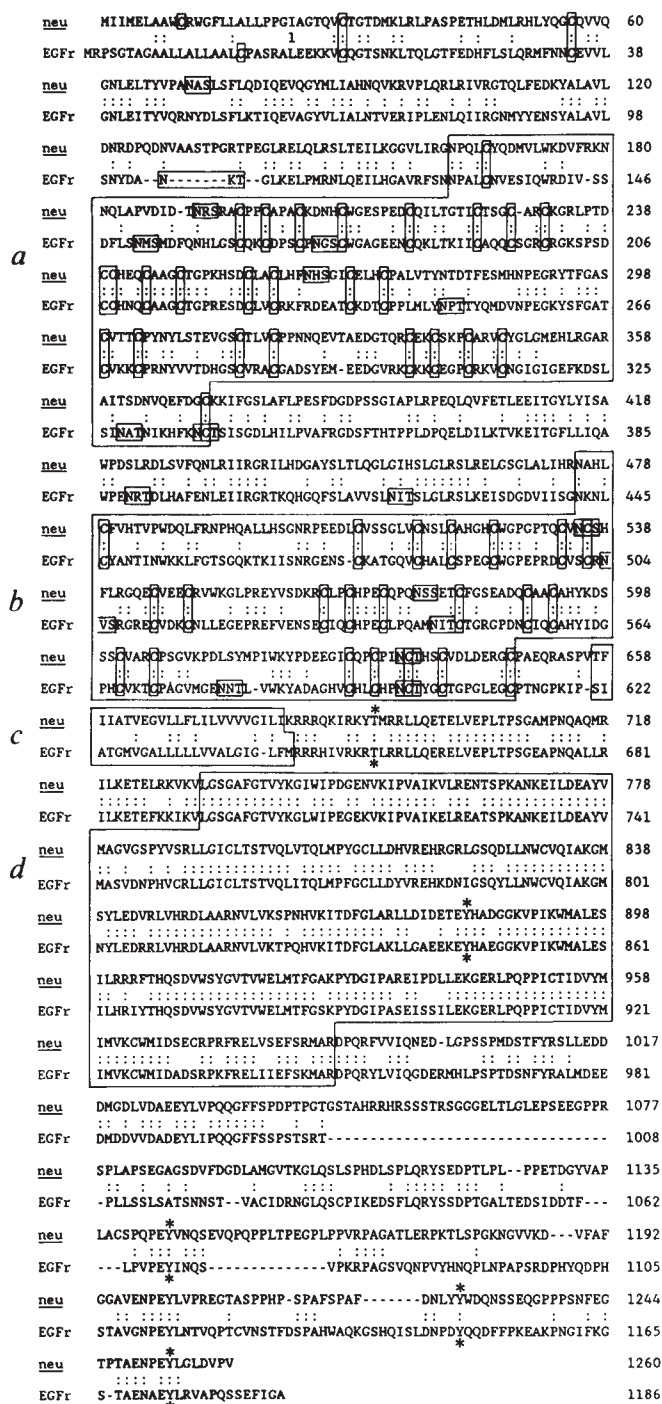


Fig. 3 Comparison of predicted amino-acid sequences of the *neu* gene product and the EGF receptor. Identities in sequence are marked by two dots between the conserved residues. All cysteine residues in the extracellular regions are boxed. Consensus sequences for *N*-linked glycosylation are marked by horizontal boxes. Asterisks denote residues which may be modified by phosphorylation. Boxed areas *a* and *b* represent the cysteine-rich domains of the extracellular part of the protein. *c*, Putative transmembrane domain; *d*, region homologous with the conserved region of the tyrosine kinase gene family. The first amino acid (leucine) of the mature EGF receptor protein is numbered (1).

an extracellular ligand-binding domain, and the 580 amino acids lying after the second hydrophobic (presumably transmembrane) domain would constitute cytoplasmic effector determinants of p185.

The putative extracellular domain of p185 contains two cysteine-rich areas between amino acids 162–372 and 475–649; these stretches contain 26 and 22 cysteines, respectively. This organization is strikingly similar to that of the comparable region of the EGF receptor¹¹. As shown in Fig. 3, the two proteins show 43% identity of amino-acid residues in the extracellular portions. Note, however, that the positions of cysteines are highly conserved. All 50 cysteines predicted in the two extracellular domains are in equivalent positions in the two proteins.

These cysteine residues are also 80–90% conserved between these proteins and the insulin receptor^{12,13}. As the peptide ligands of the insulin receptor and the EGF receptor are clearly unrelated, and as p185 does not bind either of these ligands (D. Stern and R.A.W., unpublished data), it seems that these residues must play a part other than forming the specific ligand-binding domain. They may constitute the structural scaffolding of the receptor or act to facilitate signal transduction after the receptor recognizes its ligand.

The two cysteine-rich regions of the EGF receptor bear homology with one another¹¹. Similarly, the two cysteine-rich areas of *neu* are homologous. Comparison of the two EGF receptor domains, the two *neu* domains and the insulin receptor domain reveals that 15 residues out of ~180 appear to be conserved in this family of receptors (Fig. 4a). These residues include 11 cysteines as well as several other invariant residues. There are six predicted *N*-linked glycosylation sites in *neu*. Only one corresponds precisely in position to a glycosylation site in the EGF receptor.

A stretch of 25 hydrophobic amino acids followed by a lysine and three arginines appears to separate the extracellular from the cytoplasmic determinants of *neu*. The most obvious characteristic of the cytoplasmic region is the stretch of 250 amino acids homologous with the conserved region present in all tyrosine kinases¹⁴. Previous evidence indicated that the p185 protein has such an associated enzyme activity (D. Stern and R.A.W., in preparation); the sequence strongly suggests that this activity is intrinsic to the p185 protein itself.

Figure 4b compares this kinase domain with that of several other known tyrosine kinases and the cyclic AMP-dependent protein kinase. *neu* is more closely related to the EGF receptor gene than to all other members of this gene family; 82% of the amino-acid residues in this region of the EGF receptor and p185 are identical. The *neu* gene encodes the residues conserved among the proteins specified by the tyrosine kinase gene family (25% of the total residues in this domain), including the tyrosine residue that acts as an autophosphorylation site in pp60^{src} (residue 882 in the *neu* protein product).

After the kinase domain there are 60 amino acids that are 55% identical between the EGF receptor and *neu*. These residues are followed by a stretch of 198 amino acids in *neu* that are only distantly related to the last 172 amino acids of the EGF receptor; even when many gaps are inserted, only 58 residues (~30%) are comparable. This region of the protein consists of 20% proline residues and 10% each of glycine and serine residues, with only 25% of the residues being hydrophobic. Such hydrophilicity is characteristic of comparable regions of the EGF and insulin receptors.

Three tyrosine residues have been identified in the EGF receptor that serve as acceptor sites for *in vitro* autophosphorylation; these three and a fourth, as yet unlocalized, tyrosine are present in EGF receptor labelled *in vivo*¹⁵. The major site of phosphorylation of the EGF receptor *in vivo* is Tyr 1,173, 14 residues from the carboxy terminus of the EGF receptor. A similar tyrosine seems to exist eight residues from the carboxy terminus of p185 (Tyr 1,253). Tyrosine residues in homologous positions with the other two EGF receptor phosphotyrosine sites

17537-10 from the NCI and by a grant from the American Business Cancer Research Foundation. M.-C.H. was supported by a postdoctoral fellowship from the Cancer Research Institute. R.A.W. is an American Cancer Society Research Professor.

Received 30 August; accepted 15 November 1985.

- Schubert, D. *et al.* *Nature* **249**, 224-227 (1974).
- Padhy, L. C. *et al.* *Cell* **28**, 865-871 (1982).
- Schechter, A. L. *et al.* *Nature* **312**, 513-516 (1984).
- Schechter, A. L. *et al.* *Science* **229**, 976-978 (1985).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning* (Cold Spring Harbor Laboratory, New York, 1982).
- Mulligan, R. C. & Berg, P. *Science* **209**, 1422-1427 (1980).
- Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-580 (1980).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
- Biggin, M. D., Gibson, T. J. & Hong, G. F. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3963-3965 (1983).
- Blobel, G. *et al.* *Symp. Soc. exp. Biol.* **33**, 9-36 (1979).
- Ullrich, A. *et al.* *Nature* **309**, 418-425 (1984).
- Ullrich, A. *et al.* *Nature* **313**, 756-761 (1985).
- Ebina, Y. *et al.* *Cell* **40**, 747-758 (1985).
- Bishop, J. M. A. *Rev. Biochem.* **52**, 301-354 (1983).
- Downward, J., Parker, P. & Waterfield, M. D. *Nature* **311**, 483-485 (1984).
- Hunter, T., Ling, N. & Cooper, J. A. *Nature* **311**, 480-483 (1984).
- Friedman, B. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **81**, 3034-3038 (1984).
- Davis, R. J. & Czech, M. P. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4080-4084 (1985).
- Hunter, T. *Trends biochem. Sci.* **10**, 275-280 (1985).
- Doolittle, R. F. *et al.* *Science* **221**, 275-277 (1983).
- Waterfield, M. D. *et al.* *Nature* **304**, 35-39 (1983).
- Downward, J. *et al.* *Nature* **307**, 521-527 (1984).
- Scherr, C. J. *et al.* *Cell* **41**, 665-676 (1985).
- Neckameyer, W. S. & Wang, L. H. *J. Virol.* **53**, 879-884 (1985).
- Yamamoto, T. *et al.* *Cell* **39**, 27-38 (1984).
- Yamamoto, T. *et al.* *Nature* **319**, 230-234 (1986).

Similarity of protein encoded by the human *c-erb-B-2* gene to epidermal growth factor receptor

Tadashi Yamamoto, Shuntaro Ikawa, Tetsu Akiyama*, Kentaro Semba, Nobuo Nomura†, Nobuyuki Miyajima, Toshiyuki Saito & Kumao Toyoshima

Institute of Medical Science, University of Tokyo, 4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan

* Meiji College of Pharmacy, 1-35-23, Nozawa, Setagaya-ku, Tokyo 154, Japan

† Nippon Medical School at Sakuragi, 1-10-19, Uenosakuragi, Taito-ku, Tokyo 110, Japan

A novel *v-erb-B*-related gene, *c-erb-B-2*, which has been identified in the human genome^{1,2}, maps to human chromosome 17 at q21 (ref. 40), and seems to encode a polypeptide with a kinase domain that is highly homologous with, but distinct from, that of the epidermal growth factor (EGF) receptor¹. The *c-erb-B-2* gene is conserved in vertebrates and it has been suggested¹ that the *neu* gene, detected in a series of rat neuro/glioblastomas³, is, in fact, the rat *c-erb-B-2* gene. Amplification of the *c-erb-B-2* gene in a salivary adenocarcinoma and a gastric cancer cell line MKN-7 suggests that its over-expression is sometimes involved in the neoplastic process. To determine the nature of the *c-erb-B-2* protein, we have now molecularly cloned complementary DNA for *c-erb-B-2* messenger RNA prepared from MKN-7 cells. Its sequence shows that the *c-erb-B-2* gene encodes a possible receptor protein and allows an analysis of the similarity of the protein to the EGF receptor and the *neu* product. As a consequence of chromosomal aberration in MKN-7 cells, a 4.6-kilobase (kb) normal transcript and a truncated 2.3-kb transcript of *c-erb-B-2* are synthesized at elevated levels. The latter transcript presumably encodes only the extracellular domain of the putative receptor.

Poly(A)⁺ RNA was prepared from MKN-7 cells, a gastric cancer cell line in which the *c-erb-B-2* gene is amplified⁴⁰ and over-expressed (see below). A cDNA library constructed from the MKN-7 mRNA was screened initially with a 440-base pair

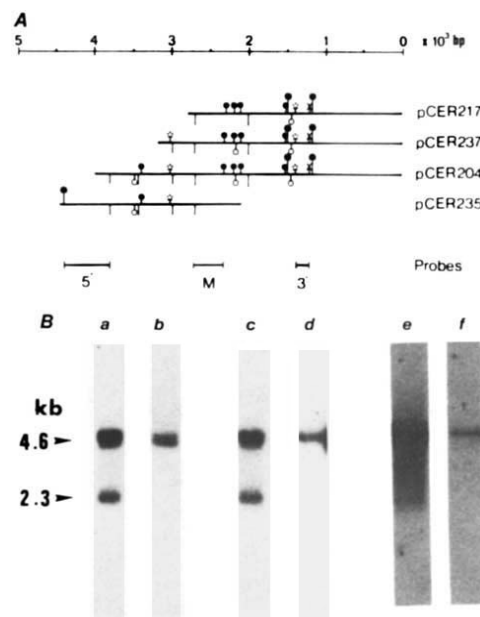


Fig. 1 Analysis of *c-erb-B-2* clones. **A**, Restriction maps of *c-erb-B-2* clones; **B**, Northern analysis of MKN-7 and placental mRNAs. **A**, Maps were constructed by the standard procedure for restriction digestion analysis of plasmid DNA⁴¹. Single or double digestion with restriction endonucleases *KpnI* (x), *EcoRI* (☆), *BamHI* (●), *PvuII* (|), *SmaI* (*) and *AccI* (○) was performed. The *AccI* site at position 2.2 kb in this map is absent from the pCER217 insert but pCER237 and pCER204 were cleaved with the enzyme at this position. Sequence data reveal that pCER217 carries a deletion of 42 bp in this region (see Fig. 2 legend). The exon sequence of the 440-bp KX DNA prepared from the genomic clone λ 107 (ref. 1) corresponds to a 150-bp *EcoRI*-*KpnI* fragment of pCER217 (3' probe). A 450-bp *PvuII*-*BamHI* fragment of pCER217 (middle probe) and a 650-bp *SmaI*-*PvuII* fragment of pCER235 (5' probe) were used for the second and third screenings, respectively, of the MKN-7 cDNA library. Total RNA was prepared from MKN-7 cells by the guanidine isothiocyanate-caesium chloride method²⁸. Poly(A)⁺ RNA was selected by two cycles of oligo(dT)-cellulose column chromatography²⁹. A cDNA library was constructed by the method of Okayama and Berg³⁰, using 5.4 μ g poly(A)⁺ RNA and 2.8 μ g vector-primer DNA. (Avian myeloblastosis virus reverse transcriptase was from Dr J. W. Beard.) *Escherichia coli* MC1061 (ref. 31) was used for transformation³². Ampicillin-resistant transformants (60,000 independent clones) were first screened by hybridization at 60 °C for 16 h in the solution described previously³³ with the 440-bp KX DNA fragment. Among 21 distinctive positive clones, pCER217 carried the longest insert of 2.8 kb. To obtain clones harbouring cDNA for further upstream sequence, 50,000 independent clones of the same library were screened with the 450-bp *PvuII*-*BamHI* fragment mapped in the 5' portion of the pCER217 insert. Of 44 positive clones, plasmid pCER237 carried the longest insert (3.2 kb). A plasmid pCER235 was assumed to be derived from an aberrant *c-erb-B-2* mRNA which shared 5' sequence with normal 4.6-kb mRNA, as described in the text. The MKN-7 cDNA library (100,000 clones) was again screened with a 650-bp *SmaI*-*PvuII* restriction fragment derived from pCER235; 48 positive clones were obtained, one of which, designated pCER204, contained the longest insert (4.0 kb) and the restriction maps of its 5' one-third and 3' two-thirds were identical to those of pCER235 and pCER217, respectively. **B**, Nitrocellulose filters containing poly(A)⁺ RNAs from MKN-7 (lanes a, c, e) and human placenta (lanes b, d, f) were hybridized with the 5' (lanes a, b), middle (lanes c, d) and 3' probes (lanes e, f) shown in **A**. Lanes b and d are photographs obtained after longer exposure of the film (overnight exposure for lanes a, c; 2 weeks exposure for lanes b, d). Poly(A)⁺ RNA (2 μ g) from MKN-7 cells was denatured with 50% formamide and 2.2 M formaldehyde and applied to a 1% agarose gel containing 2.2 M formaldehyde³⁴. RNAs on the gel were transferred to a nitrocellulose filter³⁵, which was then hybridized for 16 h under stringent conditions (50% formamide, 4 $\times$ SSC, 42 °C) with the DNA probes. After hybridization, the filter was washed under stringent conditions as described elsewhere¹. Under the hybridization conditions used, no hybridization of the DNA probes with the EGF receptor mRNA was observed. The DNA probes were labelled with [α -³²P]dCTP (3,000 Ci mmol⁻¹; Amersham) by nick-translation³⁶.

(bp) KX DNA fragment prepared from a *c-erb-B-2* genomic clone and then with cDNA probes as described in Fig. 1 legend. Among 113 positive clones, pCER204 carried the longest insert of 4.0 kb, whereas the *c-erb-B-2* mRNA is 4.6 kb long. Restriction mapping of the positive clones having an insert of >1.5 kb led to the identification of two distinct classes of clones (Fig. 1A),

Fig. 2 The *c-erb-B-2* cDNA nucleotide sequence and predicted amino-acid sequence. Nucleotides are numbered at both sides. Amino-acid sequence is numbered from the putative signal peptide above the sequence. Black dotted bars indicate the putative transmembrane region. The AATAAA box is followed (14 bp downstream) by the polyadenylated 3' end of the mRNA. The nucleotide sequence from residues 1 to 1,810 was from pCER235 and that from residues 1,583 to the extreme 3' end was from pCER217, except that the sequence of the 77-bp *Bam*HI fragment (2,314-2,390) was from pCER237, as a sequence of 42 bp was apparently deleted in pCER217 (see Fig. 1 legend). The overlapping nucleotide sequences (1,583-1,810) of pCER217 and pCER235 match. The extreme 3' sequence of pCER235, which was derived from a sequence of unknown origin caused by chromosomal translocation is not shown here. The nucleotide sequence was determined by the Maxam-Gilbert³⁷ procedure and the dideoxy chain termination method in conjunction with bacteriophage M13mp19 (refs 38, 39).

suggesting that there are two species of *c-erb-B-2* mRNA in MKN-7 cells. As cDNA clones represented by pCER235 have inserts of ≤ 2.3 kb, the other class of clones represented by pCER204 (or 217 and 237) was thought to be derived from the 4.6-kb mRNA. The pCER235 insert has a sequence that corresponds to the 5' half, but not the 3' half, of the pCER204 insert, suggesting that pCER235 was derived from mRNA which shares its sequence with the 5' portion of normal 4.6-kb mRNA. This

possibility was demonstrated by Northern hybridization of MKN-7 RNA (Fig. 1B). Poly(A)⁺ RNA from MKN-7 cells was probed with DNAs specific for different portions of the cDNA clones. Using the 3' probe, equivalent to the KX DNA probe¹, we observed a single species of 4.6-kb mRNA, which is expressed at an elevated level in MKN-7 cells (50-fold relative to placenta). However, both the *Pvu*II-*Bam*HI fragment (middle probe) and the *Sma*I-*Pvu*II fragment (5' probe) reacted with a 2.3-kb

©1986 Nature Publishing Group

Fig. 3 Amino-acid sequence of *c-erb-B-2*. **A**, Alignment of the amino-acid sequences of *c-erb-B-2* and EGF receptor (EGFR). Amino acids of the two proteins are numbered on the left. Identities in the sequences are marked by two dots between the two lines; predicted transmembrane regions are represented by dotted black bars; the possible *N*-linked glycosylation sites by wavy lines; horizontal lines indicate signal peptides (putative for *c-erb-B-2*); stars indicate cysteine residues in the sequence: solid stars are common to the two proteins and the open stars are specific to *c-erb-B-2*. The major sites of threonine and tyrosine phosphorylation of the EGF receptor or pp60^{src} are conserved in *c-erb-B-2* and are shown by open and closed triangles, respectively. The amino-acid residues corresponding to the extreme carboxy terminus of pp60^{src} and *v-erb-B* protein are indicated by vertical arrows at positions 986 and 1,228, respectively. Boxed sequences show the cysteine clusters. **B**, Schematic illustration of the *c-erb-B-2* protein. Amino-acid sequence homologies of the *c-erb-B-2* protein and the EGF receptor are shown. T, threonine and Y, tyrosine are possible phosphorylation sites. C, cysteine clusters (also boxed). —◇, Possible glycosylation sites. An open box at the amino terminus shows the putative signal peptide and that in the middle shows possible transmembrane sequence. The hydrophaticity of the *c-erb-B-2* sequence is also shown.

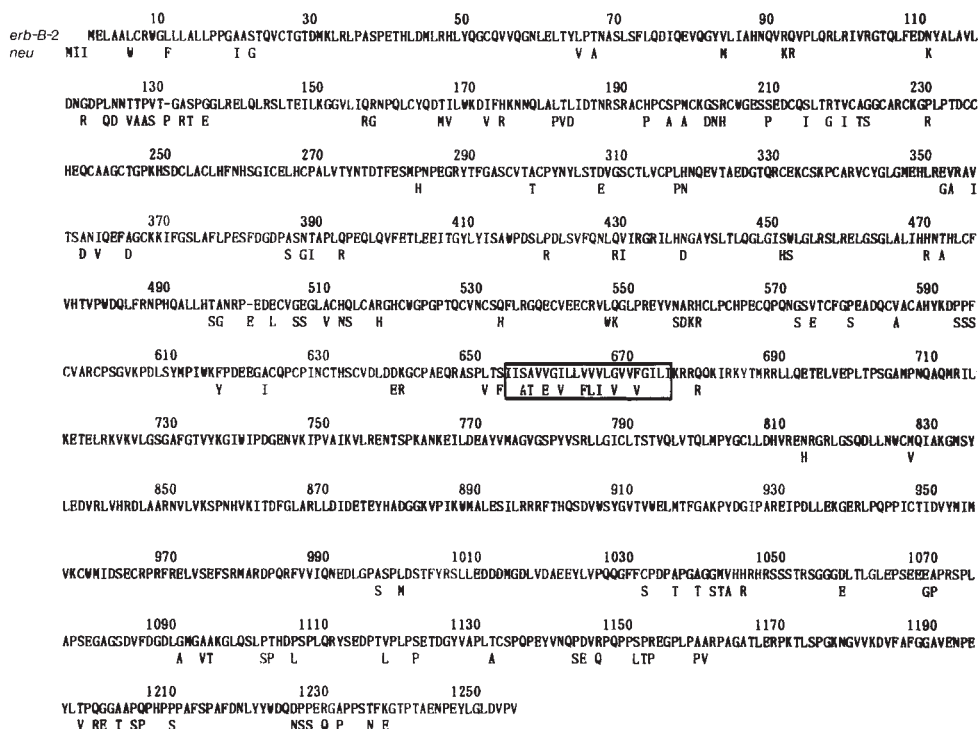


Fig. 4 Comparison of rat *neu* and human *c-erb-B-2*. The amino-acid sequence of *c-erb-B-2* is compared with that of *neu* (kindly communicated by Dr R. A. Weinberg). Only non-identical amino acids are shown for *neu*. The amino acids for *c-erb-B-2* are numbered above the sequence. The boxed sequence indicates the transmembrane domain.

could serve as a membrane-anchoring domain. This sequence, like those of other receptors⁷⁻¹⁰, is followed immediately by basic amino acids (Lys-Arg-Arg), which helps in the correct allocation of the protein at the cell surface. The first 21 amino-acid residues are also highly hydrophobic, which suggests that they represent a signal sequence for membrane glycoprotein. Eight possible sites of *N*-linked glycosylation were identified in the amino-terminal moiety of the *c-erb-B-2* protein. These data indicate that the two proteins with similar sizes have transmembrane topologies that resemble each other. Therefore, we tentatively conclude that the *c-erb-B-2* protein is a receptor for an unknown growth factor.

The sequence of the putative extracellular domain shows 44% homology with the ligand binding domain of the EGF receptor. A striking similarity is the presence of two cysteine-rich regions, in which the spatial distribution of cysteine residues is virtually identical with that in the EGF receptor. The sequences of cysteine clusters are rather hydrophilic (Fig. 3B) and would facilitate the generation of a specific conformation for signal transmission through intramolecular or intermolecular S-S bridges, as is suggested for other receptors for growth factors⁸⁻¹⁰. These findings also suggest that the extracellular domains of the *c-erb-B-2* protein and EGF receptor form similar configurations and bind to structurally related ligands. However, not only EGF but ligands such as tumour growth factor (TGF)- β , TGF- γ , fibroblast growth factor, erythropoietin, nerve growth factor, insulin and platelet-derived growth factor all failed to activate the kinase activity that is presumably intrinsic to the *c-erb-B-2* protein (data not shown).

The sequence of 260 amino acids (residues 727-986), including the postulated ATP-binding site¹¹ of the cytoplasmic domain of the *c-erb-B-2* protein, is homologous with the kinase domain of the EGF receptor (82% homology) and with retroviral oncogene products of the *src* family (25-40% homology¹). Therefore, the *c-erb-B-2* protein seems to have tyrosine kinase activity. Preliminary experiments using antiserum raised against a synthetic peptide of 14 amino-acid residues at the carboxy terminus of *c-erb-B-2* show that the *c-erb-B-2* protein gp185 exhibits tyrosine kinase activity (T.A. *et al.*, in preparation). The homology between the two proteins decreases to 32% in the

269 amino-acid residues at the carboxyl end. However, three major *in vitro* phosphorylation sites of the EGF receptor¹² are also conserved (tyrosine residues at 1,139, 1,222 and 1,248 of *c-erb-B-2*), indicating that the *c-erb-B-2* protein could also be autophosphorylated.

Another interesting feature of the *c-erb-B-2* protein is the presence of a threonine residue at position 686, surrounded by basic amino-acid residues, which is equivalent to threonine 654 of the EGF receptor on which protein kinase C-mediated phosphorylation occurs¹³. Therefore, the *c-erb-B-2* protein may be phosphorylated by protein kinase C, which would play an important role in signal transmission as suggested for the EGF receptor.

Comparison of the nucleotide sequences and deduced amino-acid sequences of the recently characterized rat *neu* (see accompanying paper¹⁴) and human *c-erb-B-2* (our present results) reveal that the *neu* gene is the rat counterpart of the *c-erb-B-2* gene (see Fig. 4 for the amino-acid sequences). Surprisingly, only two amino acids in the kinase domain (residues 813 and 817) differ between the two proteins, although we do not know whether the *c-erb-B-2* gene of MKN-7 cells can transform NIH 3T3 cells. The possible glycosylation sites of the *neu* product are located at positions corresponding to those of the *c-erb-B-2* product, except that no sequence corresponding to Asn-Asn-Thr-Thr (positions 124-127 of *c-erb-B-2*) is found in *neu*, suggesting that the mature *c-erb-B-2* protein is as large as the *neu* gene product gp185.

Evidence is accumulating that amplification and over-expression of a proto-oncogene can cause cell transformation *in vitro* and can play a part in the neoplastic process of human tumours¹⁵⁻²⁰. Amplification and elevated expression of the EGF receptor gene have been observed in glioblastomas and in squamous carcinoma cell lines²¹⁻²⁴. In contrast, amplification of the *c-erb-B-2* gene has been seen in three human adenocarcinomas: one salivary adenocarcinoma¹, one mammary carcinoma² and one MKN-7 gastric cancer cell line, suggesting that increased expression of the *c-erb-B-2* gene provides a selective advantage in the formation or the progress of tumours of epithelial cells.

Because the cDNA clone pCER235 derived from the 2.3-kb

mRNA does not contain a signal for membrane anchoring of the predicted polypeptide (data not shown), the translation product of the 2.3-kb mRNA, corresponding to the sequence of the *c-erb-B-2* ligand-binding domain, should be secreted by MKN-7 cells. Similarly, a truncated EGF receptor sequence has been reported to be over-expressed concomitantly with the EGF receptor in A431 cells^{7,25,26}, as a consequence of a strong aberration of the chromosomes²⁷. Production of the truncated receptor in MKN-7 cells may also be caused by chromosomal aberration⁴⁰. We cannot exclude the possibility that extracellular accumulation of truncated growth factor receptors is associated with the appearance of the transformed phenotype. It is also possible that they provide a growth advantage to the cells in culture, in which the growth factor receptors are over-expressed.

We thank N. Kitamura, H. Okubo and S. Nakanishi for help in constructing the MKN-7 cDNA library, H. Kawano for technical assistance and S. Sasaki for help in preparing the manuscript. We also thank Y. Kaziro for critical reading of the manuscript.

Received 28 August; accepted 27 November 1985.

1. Semba, K., Kamata, N., Toyoshima, K. & Yamamoto, T. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6497-6501 (1985).
2. King, C. R., Kraus, M. H. & Aaronson, S. A. *Science* **229**, 974-976 (1985).
3. Schubert, D. *et al. Nature* **249**, 224-227 (1974).
4. Kozak, M. *Nucleic Acids Res.* **9**, 5233-5252 (1981).
5. Kyte, J. & Doolittle, R. F. *J. molec. Biol.* **157**, 105-132 (1982).
6. Chou, P. Y. & Fasman, G. D. *A. Rev. Biochem.* **47**, 251-276 (1978).

7. Ullrich, A. *et al. Nature* **309**, 418-425 (1984).
8. Ebina, Y. *et al. Cell* **40**, 747-758 (1985).
9. Ullrich, A. *et al. Nature* **313**, 756-761 (1985).
10. Yamamoto, T. *et al. Cell* **39**, 27-38 (1984).
11. Kamps, M. P., Taylor, S. S. & Sefton, B. M. *Nature* **310**, 589-592 (1984).
12. Downward, J., Parker, P. & Waterfield, M. D. *Nature* **311**, 483-485 (1984).
13. Hunter, T., Ling, N. & Cooper, J. A. *Nature* **311**, 480-483 (1984).
14. Bargmann, C. I., Hung, M.-C. & Weinberg, R. A. *Nature* **319**, 226-230 (1986).
15. Chang, E. H., Furth, M. E., Scolnick, E. M. & Lowy, D. R. *Nature* **297**, 479-483 (1982).
16. Gazit, A. *et al. Cell* **39**, 89-97 (1984).
17. Clarke, M. F. *et al. Nature* **308**, 464-467 (1984).
18. Schwab, M. *et al. Nature* **305**, 245-248 (1983).
19. Kohl, N. E. *et al. Cell* **35**, 359-367 (1983).
20. Schwab, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 4940-4944 (1984).
21. Libermann, T. A. *et al. Cancer Res.* **44**, 753-760 (1984).
22. Libermann, T. A. *et al. Nature* **313**, 144-147 (1984).
23. Cowley, G., Smith, J. A., Gusterson, B., Hendler, F. & Ozanne, B. in *Cancer Cell 1* (eds Levine, A. J., VandeWoude, G. F., Topp, W. C. & Watson, J. D.) 5-10 (Cold Spring Harbor Laboratory, New York, 1984).
24. Yamamoto, T. *et al. Cancer Res.* (in the press).
25. Merlino, G. T. *et al. Science* **224**, 417-419 (1984).
26. Weber, W., Gill, G. N. & Spiess, J. *Science* **224**, 294-297 (1984).
27. Merlino, G. T. *et al. Molec. cell. Biol.* **5**, 1722-1734 (1985).
28. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Ruffer, W. J. *Biochemistry* **18**, 5294-5299 (1979).
29. Aviv, H. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **69**, 1408-1412 (1972).
30. Okayama, H. & Berg, P. *Molec. cell. Biol.* **2**, 161-170 (1982).
31. Casadaban, M. & Cohen, S. N. *J. molec. Biol.* **138**, 179-207 (1980).
32. Hanahan, D. & Meselson, M. *Meth. Enzym.* **100**, 333-342 (1983).
33. Iwai, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 7405-7409 (1983).
34. Lebrach, H., Diamond, K., Wozney, J. M. & Boedtker, H. *Biochemistry* **16**, 4743-4751 (1977).
35. Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201-5205 (1980).
36. Maniatis, T., Jeffery, A. & Kleid, D. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1184-1188 (1975).
37. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 109-111 (1979).
38. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
39. Messing, J. *Meth. Enzym.* **101**, 20-77 (1983).
40. Fukushige, S. *et al. Molec. cell. Biol.* (in the press).
41. Maniatis, T., Fritsch, E. & Sambrook, J. in *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).

Primate η -globin DNA sequences and man's place among the great apes

Ben F. Koop*, Morris Goodman*§, Peilin Xu†, Keith Chan† & Jerry L. Slightom†‡

*Department of Anatomy, Wayne State University School of Medicine, Detroit, Michigan 48201, USA

†Laboratory of Genetics, University of Wisconsin, Madison, Wisconsin 53706, USA

‡Upjohn Company, Kalamazoo, Michigan 49001, USA

Molecular studies indicate that chimpanzee and gorilla are the closest relatives of man (refs 1-7 and refs therein). The small molecular distances found point to late ancestral separations^{2,4,7}, with the most recent being between chimpanzee and man, as judged by DNA hybridization^{3,8}. Kluge⁹ and Schwartz¹⁰ contest these conclusions: morphological characters group a chimpanzee-gorilla clade with the Asian ape orang-utan in Kluge's cladistic study and with an orang-utan-human clade in Schwartz's study. Clearly, extensive sequencing of nuclear DNA is needed to resolve by cladistic analysis the branching order within Hominoidea¹¹. Towards this goal, we are sequencing orthologues of the primate $\psi\eta$ -globin locus^{12,13}. Here, we compare the newly completed sequences of orang-utan and rhesus monkey with human, chimpanzee, gorilla, owl monkey, lemur and goat orthologues. Our findings substantially increase the evidence indicative of a human-chimpanzee-gorilla clade with ancestral separations around 8 to 6 Myr ago. We also verify that neutral hominoid DNA evolved at markedly retarded rates.

The η locus is one of five ancient β -related globin genes linked in a cluster 5'- ϵ - γ - δ - β -3' that arose from tandem duplications (200-100 Myr)^{12,13}. This ancient η gene was embryonically expressed in early eutherians and persisted as a functional gene in artiodactyls, but became a pseudogene in proto-primates and was lost from rodents and lagomorphs. Previous work from this laboratory¹³ established that the goat η gene sequence

Fig. 1 (Opposite) Aligned nucleotide sequences of seven primate $\psi\eta$ -globin genes and the active goat η -globin gene. Total DNA was isolated from orang-utan (*Pongo pygmaeus*) no. 1 liver (Yerkes Primate Center) and from a rhesus monkey (*Macaca mulatta*) blood sample (California Primate Center). DNAs were subjected to a series of limited *EcoRI* digestions and size-selected fragments, 15-25 kilobases (kb), were cloned into the λ vector Charon 32³³ by the procedure described by Slightom *et al.*³⁴. Recombinant phage DNAs were packaged into phage capsids using the *in vitro* phage packaging procedure³⁵. Charon 32 phage were plated on the *recA*⁻ *Escherichia coli* host ED8767³⁶ and screened using a ³²P-labelled 245-base-pair (bp) *AvaII*-*EcoRI* fragment isolated from the γ -globin cDNA clone pJW151³⁷. The $\psi\eta$ -containing *EcoRI* fragments were subcloned into pBR322 (orang 7.0 kb, rhesus 10.0 kb). DNA sequencing was done using the chemical procedure described by Maxam and Gilbert³⁸. The nucleotide sequences for human, chimpanzee and gorilla are from Chang and Slightom¹⁶, and owl monkey and lemur are from Harris *et al.*¹². Only the 5'-flanking to exon 2 of the lemur sequence is orthologous to the η gene^{12,15}; therefore only this part of the sequence is used in our analyses. The goat η -globin gene (goat ϵ ¹¹ gene) nucleotide sequence is from Shapiro *et al.*¹⁴. The nucleotide sequencing numbering system is based on the overall alignment among these sequences. The complete nucleotide sequence for the human $\psi\eta$ -globin gene (HUMA) is presented on the top line and differences are given for the remaining sequences. Asterisks indicate the presence of gaps placed to minimize the number of geneic changes during descent^{13,39}. The simian $\psi\eta$ genes are divided into three exons separated by two introns each obeying the GT-AG splicing rule⁴⁰. Lemur $\psi\eta$ ^{12,15} extends only to the 3' end of exon 2 and has a defective intron 1 splice site. Intron 1 sequences are all 121 bp in length, while intron 2 lengths vary between 841 to 877 bp. The 5'-flanking region of the orang $\psi\eta$ gene contains a 38-bp direct repeat (positions 205-244 and 245-282) common to hominoid $\psi\eta$ genes. Each repeat contains a CCAAT type promoter element. The RNA polymerase II binding site (TATAA) has diverged in the $\psi\eta$ genes from that found in normal β -type globin genes⁴¹ (note positions 299-304). As in the other primate $\psi\eta$ genes the initiation codon (INT positions 384-386) of orang and rhesus show defects that prohibit translation. All of the primate $\psi\eta$ genes share a terminator sequence TGA (positions 1,885-1,888) and have the canonical poly(A) addition signal⁴². With regard to the direct repeat at positions 1,895-1,914 and 1,915-1,940, the 22-base gap (positions 1,895-1,914) in orang and rhesus could be the result of either two independent deletions or a deletion of the 5' direct repeat in the stem catarrhines followed by a reduplication in the stem hominines. Exons terminate at the ---> sites.

§To whom correspondence should be addressed.

[illegible]

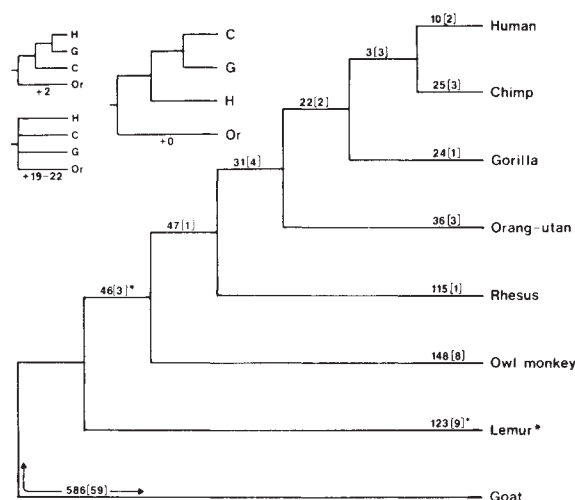


Fig. 2 Species relationship as indicated by a parsimony analysis^{43,44} of the aligned sequences in Fig. 1. The upper right tree required 1,075 mutation events (1,218 after correction^{51,32} for multiple substitutions occurring at the same site). Alternative trees (upper left) required 0, +2, +19–22 additional events. The 12 trees requiring +19–22 additional events resulted from breaking up a monophyletic African ape–human clade by introduction of orang. Branch lengths include both the corrected number of substitutions and the number of insertion/deletion events (the number of insertion/deletion events alone are given in brackets). The asterisk indicates the branches and corresponding link lengths in which geneic events were recorded only over those alignment positions that the much shorter lemur $\psi\eta$ sequence shared with other sequences. Normalization of the starred values requires multiplication by approximately 2.6. Ancestral sequences for the nodes of the upper right tree were deduced by our maximum parsimony method; from 548 ancestral-to-descendant substitutions occurring within the primate $\psi\eta$ gene region, the calculated⁴⁵ relative frequencies of the 12 types of substitutions were 20.5% G→A, 20.3% C→T, 14.3% A→G, 12.3% T→C, 5.9% G→C, 5.7% G→T, 4.7% C→A, 4.0% T→G, 3.3% A→T, 3.1% C→G, 2.8% T→A. The expected base composition at equilibrium (19% G, 29% A, 31% T and 21% C), calculated^{46,47} from the above relative frequencies, was found to approximate to that of the large introns of β -globin-related genes (21% G, 29% A, 33% T and 17% C, averaged over 33 such introns) which are thought to represent selectively neutral DNA. Within the primate $\psi\eta$ gene regions the distribution of relative frequencies of substitutions was not found to differ significantly between exon and non-exon regions, nor did the exon and non-exon regions differ in overall substitution rates. Compared with the base composition of the exons of goat η and other functional globin genes (28% G, 22% A, 23% T and 27% C, averaged over 72 globin genes), the base composition of primate $\psi\eta$ exons (26% G, 26% A, 25% T and 23% C) is clearly evolving towards the equilibrium composition of neutral DNA.

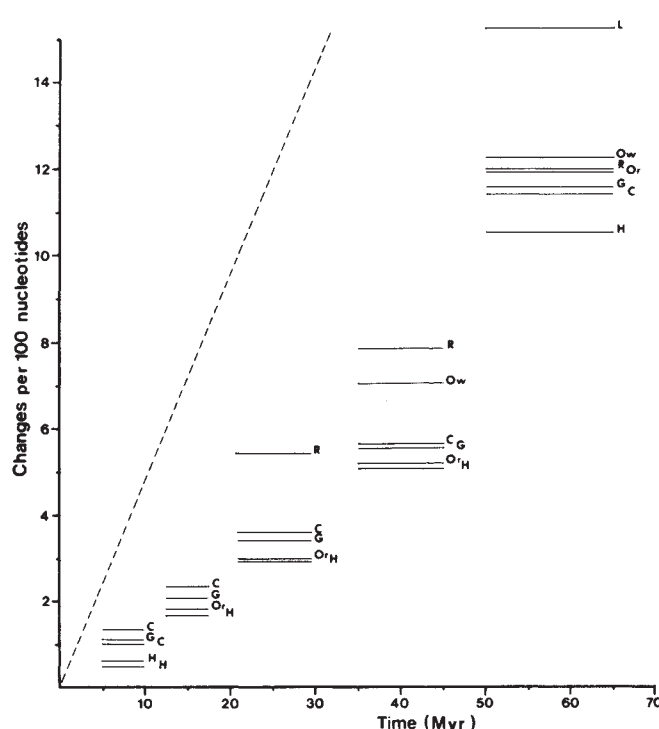


Fig. 3 The number of changes per 100 nucleotides from ancestral nodes to extant species plotted against a range of palaeontological time estimates^{11,23–28}. The dashed line represents 5×10^{-9} substitutions per site per yr. 50–65 Myr is suggested as the divergence time of the prosimian and simian lineages, 35–45 Myr indicates Platyrrhini–Catarrhini divergence, 20–30 Myr indicates cercopithecoid–hominoid divergence, 13–18 Myr indicates *Pongo*–Hominine divergence and 5–10 Myr indicates human–chimpanzee–gorilla divergence plus human–chimpanzee (chimpanzee–gorilla) divergence times. The following labels for species are used: L, lemur; Ow, owl monkey; R, rhesus; Or, orang-utan; G, gorilla; C, chimpanzee; H, human. The slope of the dashed line represents the substitution values determined for mammalian pseudogenes other than the $\psi\eta$ genes and has been equated to the neutral mutation rate^{18,19}.

(originally named ϵ^{11})¹⁴ and primate η pseudogene sequences (each originally named either $\psi\beta^{15,16}$ or $\psi\beta^{12,17}$) are orthologues. With addition of rhesus and orang orthologues (shown in Fig. 1 aligned against other η sequences), the seven primate pseudogenes represent most major taxonomic groups within primates. The rhesus and orang sequences prove to be typical representatives of this $\psi\eta$ locus; they contain mutations which inactivate functional genes (defective initiation codon at position 384–386, a stop codon at position 429–431, and frame-shift deletions at orang position 447 and rhesus positions 441, 447 and 1,803; Fig. 1). The defective A→G substitution in the initiation codon is shared by all primate sequences, thereby supporting the deduction that the η -globin locus became non-functional in proto-primates to the primate radiation (approximately 60 Myr).

Trees describing the genealogical relationships of the η -globin sequences were constructed by the most parsimonious placement of nucleotide substitutions. This cladistic procedure does not require clock-like behaviour in rates of change but only that matches in nucleotide sequence among taxa result more often from shared common ancestry than from convergencies, reversals or parallelisms. The two most parsimonious trees found (Fig. 2) required 1,075 independent genetic events to explain the pattern of variation among the eight η -globin sequences. One of these two trees, after separation of *Gorilla*, places only three derived mutations on the common stem of a *Homo*(human)–*Pan*(chimpanzee) branch: a deletion at position 1,491–1,494, an insertion at position 1,739 and a deletion at position 1,916. The last event is equivocal because an alternative parsimony solution redistributes the duplication and deletion events that occur in that region (positions 1,895–1,941). In the other most parsimonious tree, *Homo* separates from the common stem of a *Pan*–*Gorilla* branch; this latter grouping is obtained by two transitional substitutions (G→A) at positions 2,213 and 2,238. No events uniquely support a human–gorilla branch. As evident from Fig. 2, a Homininae branch of *Homo*–*Pan*–*Gorilla* is strongly supported; there are 22 derived mutations on its common stem: an insertion at positions 1,385–1,391, a deletion at position 1,879 and 20 nucleotide substitutions of which 17 are unequivocal (positions 121, 300, 326, 422, 631, 654, 750, 1,185, 1,214, 1,263, 1,307, 1,381, 1,467, 1,484, 1,592, 1,677, 1,840).

Table 1 Inter-taxonomic DNA divergences from DNA hybridization and η nucleotide sequence data

Species comparisons	DNA hybrids		ETA gene	
	TMH	TMH-C	Dist.	Dist.-C
Huma.-Chim.	1.9*	1.9	1.7	1.7
Huma.-Gori.	2.4*	2.4	1.7	1.7
Huma.-Oran.	3.6*	3.7	3.4	3.4
Chim.-Gori.	2.2*	2.2	2.2	2.3
Chim.-Oran.	3.7*	3.8	4.0	4.1
Gori.-Oran.	3.8*	3.9	3.9	4.0
Homi.-Cerc.	7.7*	8.1	8.0	8.5
Homi.-Cebo.	12.2†‡	13.3	11.6	12.6
Homi.-Pros.	22.8†‡	27.2	23.9	28.8
Homi.-Euth.	31.6‡§	41.0	30.9	39.8
Cerc.-Cebo.	13.5†	14.9	12.9	14.1
Cerc.-Pros.	23.5†¶	28.2	24.3	29.4
Cerc.-Euth.	—	—	32.6	42.9
Cebo.-Pros.	24.0†¶	28.9	25.5	31.1
Cebo.-Euth.	—	—	32.8	43.1
Pros.-Euth.	29.3‡§	37.2	27.4	34.1

TMH indicates median sequence divergence determined from thermal stability measurements^{8,29,30}. Dist. indicates divergence determined from nucleotide sequences. TMH-C or Dist.-C indicates divergence after correction for multiple substitutions^{31,32}. For taxonomic comparisons the following abbreviations are used: Huma., human; Chim., chimpanzee; Oran., orang-utan; Homi., homininae; Cerc. Cercopithecoidea; Cebo., Ceboidea; Pros., lemurs (or galago); Euth., goat (or tree shrew). Note that comparisons involving the goat η sequence are underestimates due to the inclusion of conserved functional exon regions in these comparisons. This is particularly true in the lemur-goat η comparison in which the shortened lemur sequence corresponds to the most conserved region (exons 1 and 2) of the functional goat η gene.

*Values from ref. 8.

†Values from ref. 29.

‡Values from ref. 30.

§Divergence from tree shrew rather than goat.

¶Divergence from galago rather than lemur.

The genealogical family Hominidae (Homininae plus *Pongo*) is also very strongly supported; it is defined by three deletions (at positions 164, 966-970, 1,610-1,637), an insertion (positions 245-282) and 27 separate substitutions of which 24 are unequivocal (positions 5, 66, 383, 405, 495, 573, 780, 853, 926, 1,268, 1,278, 1,334, 1,422, 1,667, 1,859, 1,949, 2,002, 2,093, 2,137, 2,138, 2,161, 2,188, 2,193, 2,213). A total of 47 shared derived mutations define infraorder Catarrhini and another 46 define suborder Anthropoidea.

Rates of $\psi\eta$ sequence evolution were calculated from the number of events on the branches of our genealogical reconstruction (Fig. 2). It is apparent from Fig. 3 that marked variations in lineage rates exist. The fastest rates are shown by lemur (2.7×10^{-9}) and in the early common simian stem (2.9×10^{-9}), intermediate rates by owl monkey and rhesus monkey and the slowest rates by hominoids (averaging 1.3×10^{-9} in descent from the catarrhine branchpoint). All primate rates are less than the average mammalian neutral rate^{18,19} and the two slowest rates, those of human and orang lineages, are only one-fifth this average neutral rate.

It follows from the close correspondence of interspecies DNA divergence estimated by DNA hybridization and $\psi\eta$ sequence results (Table 1) that the retardation of primate rates evident in the $\psi\eta$ sequences applies to total single-copy DNA. This supports Britten's analysis²⁰, which demonstrated that rates of neutral DNA sequence evolution vary by as much as a factor of five, with hominoids and some bird taxa showing the slowest rates and rodents, *Drosophila* and sea urchins showing the fastest rates. Explanations for the retarded hominoid rate invoke lengthening generation times, improved DNA repair mechanisms and combinations of both factors²⁰⁻²².

Clearly varying rates of evolution in different lineages affect the ability of a global molecular clock to predict specific divergence times. An alternative approach is to focus on a more localized set of branchpoints within a particular lineage and to use one or more well-established palaeontological time points within that lineage to calibrate a local clock. We have carried out such local clock calculations for Catarrhini, Hominidae and Homininae using the Anthropoidea branchpoint as an outside reference on assigning a palaeontological time of 35-45 Myr^{23,24} to this reference. The date found for the Catarrhini branchpoint is 22.2-28.2 Myr, the Hominidae (*Pongo*-Homininae) branchpoint is 12.7-16.4 Myr and the Homininae branchpoint is 6.3-8.1 Myr. These local clock dates correspond well with assessments of the hominoid fossil record^{11,25-28}. However, applying the local clock calibration to simian-lemur, primate-non-primate and *Mus*-*Rattus* DNA hybridization distances results in dates that are too ancient (72-92 Myr, 98-125 Myr and 52-68 Myr).

In conclusion, to a high degree of certainty, the DNA sequence data have established the genealogical validity of grouping *Homo*, *Pan* and *Gorilla* and excluding *Pongo* in a monophyletic subfamily Homininae. However, within Homininae the relationships remain unclear, perhaps due to relative close divergence points. Our analysis of the primate $\psi\eta$ globin genealogy also raises questions concerning the validity of using global clocks to estimate divergence times. Because of variability in rates of divergence in different lineages, we suggest that local clocks in some cases yield more accurate divergence times.

We thank Dr E. A. Zimmer for the gift of orang-utan No. 1 DNA, Dr S. Martin for rhesus DNA, J. Renke and J. Bliska for help in isolating phage and Dr N. Maeda for communicating restriction enzyme map data of the rhesus β -globin gene cluster before publication. J.L.S. thanks Drs Frederick R. Blattner and Oliver Smithies for the use of shared laboratory space and equipment in the Department of Genetics, University of Wisconsin. We also thank Wen-Hsiung Li, Roy J. Britten and Raoul E. Benveniste for providing us with pre-publication manuscripts. This study was supported by grants from NIH, HD16595 (J.L.S.), NSF, BSR82-07336 (M.G.) and Alfred P. Sloan Foundation (M.G.).

Received 12 August; accepted 13 November 1985.

- Goodman, M. *Ann. N.Y. Acad. Sci.* **102**, 219-234 (1962); in *Classification and Human Evolution* (ed. Washburn, S. L.) 204-234 (Aldine, Chicago, 1963).
- Sarich, V. M. & Wilson, A. C. *Science* **158**, 1200-1203 (1967).
- Hoyer, B. H., van de Velde, N. W., Goodman, M. & Roberts, R. B. *J. hum. Evol.* **1**, 645-649 (1972).
- Sarich, V. M. & Cronin, J. E. in *Molecular Anthropology* (eds Goodman, M. & Tashian, R. E.) 141-170 (Plenum, New York, 1976).
- Brown, W. M., Prager, E. M., Wong, A. & Wilson, A. C. *J. molec. Evol.* **18**, 225-239 (1982).
- Tashian, R. E., Hewett-Emmett, D. & Goodman, M. in *Isozymes* Vol. 7 (eds Rattozzi, M. C., Sandalias, J. G., Siciliano, M. Y. & Whitt, G. S.) 79-100 (Liss, New York, 1983).
- Goodman, M., Braunitzer, G., Stangl, A. & Schrank, B. *Nature* **303**, 546-548 (1983).
- Sibley, C. G. & Ahlquist, J. E. *J. molec. Evol.* **20**, 2-15 (1984).
- Kluge, A. G. in *New Interpretations of Ape and Human Ancestry* (eds Ciochon, R. L. & Corruccini, R. S.) 151-177 (Plenum, New York, 1983).
- Schwartz, J. H. *Nature* **308**, 501-505 (1984).
- Andrews, P. *Nature* **314**, 498-499 (1985).
- Harris, S., Barrie, P. A., Weiss, M. L. & Jeffreys, A. J. *J. molec. Biol.* **180**, 784-802 (1984).
- Goodman, M., Koop, B. F., Czelusniak, J., Weiss, M. L. & Slightom, J. L. *J. molec. Biol.* **180**, 803-823 (1984).
- Shapiro, S. G., Schon, E. A., Townes, T. M. & Lingrel, J. B. *J. molec. Biol.* **169**, 31-52 (1983).
- Jeffreys, A. J. *et al. J. molec. Biol.* **156**, 487-503 (1982).
- Chang, L.-Y. E. & Slightom, J. L. *J. molec. Biol.* **180**, 767-783 (1984).
- Jagadeeswaran, P., Pan, J., Forget, B. G. & Weisman, S. M. *Cold Spring Harb. Symp. quant. Biol.* **47**, 1079-1086 (1983).
- Hayashida, H. & Miyota, T. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2671-2675 (1983).
- Li, W.-H., Luo, C.-C. & Wu, C.-I. in *Molecular Evolutionary Genetics* (ed. MacIntyre, R.) (Plenum, New York, in the press).
- Britten, R. J. *Science* (in the press).
- Neel, J. V. *J. Heredity* **74**, 2-15 (1983).
- Goodman, M. *BioEssays* **3**, 9-14 (1985).
- Rosenberger, A. L. *J. hum. Evol.* **13**, 737-742 (1984).
- Gingerich, P. D. *Ybk phys. Anthropol.* **27**, 57-72 (1984).
- Andrews, P. J. in *New Interpretations of Ape and Human Ancestry* (eds Ciochon, R. L. & Corruccini, R. S.) 440-463 (Plenum, New York, 1983).
- Ciochon, R. L. in *New Interpretations of Ape and Human Ancestry* (eds Ciochon, R. L. & Corruccini, R. S.) 783-843 (Plenum, New York, 1983).
- Pitbeam, D. *Scient. Am.* **250** (3), 84-97 (1984).
- Hill, A. *Nature* **315**, 222-224 (1985).

29. Benveniste, R. E. in *Molecular Evolutionary Genetics* (ed. MacIntyre, R.) (Plenum, New York, in the press).
30. Bonner, T. I., Heinemann, R. & Todaro, G. J. *Nature* **286**, 420-423 (1980).
31. Jukes, T. H. & Cantor, C. R. in *Mammalian Protein Metabolism* (ed. Munro, H. N.) 21-123 (Academic, New York, 1969).
32. Gojobori, T., Ishii, K. & Nei, M. *J. molec. Evol.* **18**, 414-423 (1982).
33. Loenen, W. A. M. & Blattner, F. R. *Gene* **26**, 171-179 (1983).
34. Slightom, J. L., Blechl, A. E. & Smithies, O. *Cell* **21**, 627-638 (1980).
35. Hohn, B. *Meth. Enzym.* **68**, 299-309 (1979).
36. Murray, N. E., Brammer, W. J. & Murray, K. *Molec. gen. Genet.* **150**, 53-61 (1977).
37. Wilson, J. T. *et al. Nucleic Acids Res.* **5**, 563-581 (1978).
38. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 400-560 (1980).
39. Smith, F. & Waterman, M. S. *J. molec. Biol.* **147**, 195-197 (1981).
40. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. & Chambon, P. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4853-4857 (1978).
41. Efstratiadis, A. *et al. Cell* **21**, 653-668 (1980).
42. Proudfoot, N. J. & Brownlee, G. G. *Nature* **263**, 211-214 (1976).
43. Goodman, M. *Prog. biophys. molec. Biol.* **38**, 105-164 (1981).
44. Czelusniak, J. *et al. Nature* **298**, 297-300 (1982).
45. Gojobori, T., Li, W.-H. & Graur, D. *J. molec. Evol.* **18**, 360-369 (1982).
46. Wright S. *Evolution and the Genetics of Populations* Vol. 2 (University of Chicago Press, 1969).
47. Tajima, F. & Nei, M. *J. molec. Evol.* **18**, 115-120 (1982).

fgr proto-oncogene mRNA induced in B lymphocytes by Epstein-Barr virus infection

Marc S. C. Cheah*, Timothy J. Ley†, Steven R. Tronick* & Keith C. Robbins*

* Laboratory of Cellular and Molecular Biology, National Cancer Institute and

† Clinical Hematology Branch, National Heart, Lung and Blood Institute, Bethesda, Maryland 20205, USA

Several acute transforming retroviruses encode tyrosine-specific protein kinases which possess structural and functional relationships to cell-surface receptors for certain growth factors^{1,2}. One such tyrosine kinase is encoded by the *onc* gene, *v-fgr*, of Gardner-Rasheed feline sarcoma virus (GR-FeSV)³⁻⁶. Recently, we have isolated and characterized the human gene, *c-fgr*, corresponding to the viral *onc* sequence and have shown that *c-fgr* is a unique gene located on the short arm of chromosome 1 (ref. 7). Here we report that certain lymphomas (but not sarcomas or carcinomas) express *fgr*-related messenger RNA. This transcript is detected in Burkitt's lymphoma cell lines naturally infected with Epstein-Barr virus (EBV), but not in EBV-negative Burkitt's lymphoma cells. Normal umbilical cord or peripheral blood lymphocyte lines established *in vitro* by EBV infection also contain detectable *c-fgr* mRNA. Moreover, a 50-fold increase of the steady-state *c-fgr* mRNA concentration is observed when uninfected Burkitt's lymphoma cell lines are deliberately infected with EBV. These findings demonstrate for the first time the induction of a proto-oncogene in response to infection by a DNA tumour virus.

To investigate whether the *fgr* proto-oncogene was expressed in human tumour cells, RNAs were prepared from selected cell lines representing a diverse spectrum of human cancers. After selection on oligo(dT) columns, poly(A)⁺ RNAs were examined for the presence of *v-fgr*-related mRNA by Northern blotting. *c-fgr* transcripts were detected in 6 of 11 cell lines derived from myelo- or lymphoproliferative disorders, but not in any of the 9 carcinoma or 6 sarcoma cell lines examined. In a representative experiment (Fig. 1) the *fgr*-related mRNA detected was 3 kilobases (kb) long, which is in close agreement with the length of the transcript detected in normal human lung tissue⁷.

Because our initial survey had indicated the presence of the *c-fgr* transcript in cell lines derived from lymphoproliferative disorders, our analysis was extended to include various B-cell lines derived from African and American undifferentiated lymphomas of the Burkitt and non-Burkitt types. As summarized in Table 1, the *c-fgr* transcript was detected in approximately half of the lymphoma cell lines tested. All cell lines infected with EBV expressed detectable *fgr* mRNA. One non-Burkitt's

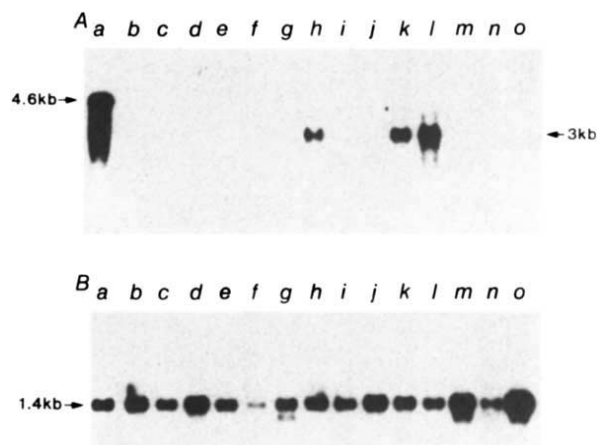


Fig. 1 A, Survey of human tumour cell lines for expression of *c-fgr*-related RNA. Poly(A)⁺-containing RNA was prepared, as described elsewhere²², from human tumour cell lines and examined by the Northern technique using a *v-fgr*-specific DNA fragment as probe⁷. Carcinoma-derived cell lines A2054, A549, A2182, A431, A1847 and A2199 (ref. 23 and S. A. Aaronson and N. Ellmore, unpublished) (lanes b-g, respectively), lymphoma-derived cell lines JI, MC116, JD38, Raji and Namalwa (lanes h-l, respectively), as well as sarcoma-derived cell lines A375, A1632 and A2984 (ref. 23 and S. A. Aaronson and N. Ellmore, unpublished) (lanes m-o, respectively), were examined. RNA from GR-FeSV-infected mink cells (lane a) was used as a control. The relative amounts of hybridizable RNA transferred to nitrocellulose filters were shown to be similar by rehybridization with a GR-FeSV-derived, actin-specific probe (B).

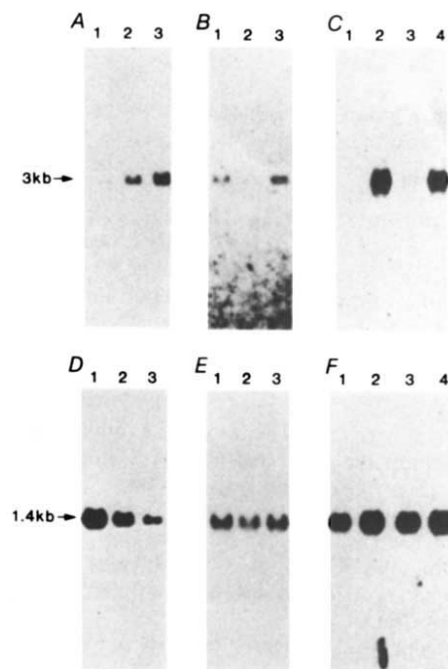


Fig. 2 Expression of *fgr*-related mRNA in cells infected with EBV *in vitro*. Poly(A)⁺-containing RNAs prepared from EBV-infected normal peripheral (A) or umbilical cord (B) blood lymphocytes, as well as uninfected (C, lanes 1, 3) or EBV-infected (C, lanes 2, 4) BJAB (lanes 1, 2) or Ramos (lanes 3, 4) Burkitt's lymphoma cell lines^{24,25}, were analysed as described in Fig. 1 legend. Relative amounts of hybridizable RNA transferred to nitrocellulose filters were shown to be similar by rehybridization with a GR-FeSV-derived actin-specific probe (D-F).

lymphoma cell line that expressed detectable *c-fgr* mRNA was not nuclear antigen (EBNA)-positive. These results revealed an association of *fgr* proto-oncogene mRNA with Burkitt's lymphoma cells which had been naturally infected with EBV.

To further assess the association between expression of the *fgr* proto-oncogene and EBV infection, we examined RNAs prepared from normal umbilical cord or peripheral blood lymphocyte cell lines established as a result of EBV infection. Figure 2 shows that the *fgr* proto-oncogene transcript was expressed at detectable levels in each of these cell lines, providing further evidence that EBV infection was responsible for the increased concentration of *fgr* proto-oncogene transcripts detected in the lymphoid cells. Preliminary efforts to determine whether normal B lymphocytes express *c-fgr* mRNA have revealed that normal peripheral blood mononuclear cells contain this transcript. However, such preparations consist of a variety of cells, of which only about 10% are B lymphocytes. Thus, further studies will be required to identify the source of *c-fgr* mRNA in peripheral blood (see note added in proof). In any case, the fact that normal lymphocytes become immortalized as stable cell lines in response to EBV infection, but do not acquire fully neoplastic properties, suggests that expression of *c-fgr* RNA alone is not sufficient to induce the neoplastic state.

We extended our analysis to EBV-negative Burkitt's lymphoma cells which are susceptible to EBV infection. BJAB cells infected with a prototype EBV, B95-8, which is capable of immortalizing normal B lymphocytes^{8,9}, or Ramos cells infected with a nontransforming derivative, P3HR-1 (ref. 10), were examined for the presence of *fgr* proto-oncogene transcripts. These converted cell lines are known to closely resemble their parental lines in morphology and with regard to karyotype and HLA- and B-antigen typing¹¹. As shown in Fig. 2C, uninfected BJAB or Ramos cell lines expressed little or no detectable *c-fgr* mRNA, whereas cells infected with either of the two EBV strains expressed the *c-fgr* transcript. We conclude from these results that the increase in *c-fgr* RNA concentration was a specific event associated with EBV infection. Moreover, as infection with either an immortalizing or a non-immortalizing EBV strain led to an increase in the steady-state concentration of *fgr* mRNA, we suggest that altered *c-fgr* expression alone is insufficient for lymphocyte immortalization.

To assess the level of *c-fgr* transcript induction in response to EBV infection, DNA fragments containing a dihydrofolate reductase exon (*dhfr*)¹² or human *c-fgr* sequences⁷ were used

as probes in quantitative *S*₁ nuclease protection experiments. As a positive control for each probe, RNA prepared from normal human peripheral blood mononuclear cells was examined (Fig. 3). As expected, a major band of 144 nucleotides was protected with the *dhfr* probe in the cell lines that were in cycle¹². The *c-fgr* DNA probe detected sequences of 165 and 120 nucleotides, sizes which were consistent with those of *v-fgr*-related stretches present in the DNA fragment used as a probe (data not shown). The intensity of bands detected with each DNA probe increased as a function of RNA concentration, making it possible to quantitate the level of *c-fgr* mRNA induction. RNAs from uninfected or EBV-infected cell lines contained similar amounts of *dhfr* mRNA, as indicated by the intensity of the 144-nucleotide *dhfr* probe fragment (Fig. 3B). In contrast, *c-fgr* sequences of 165 and 120 nucleotides were protected by RNA from EBV-infected but not uninfected BJAB or Ramos cells. By varying the exposure times, we calculated that the steady-state concentration of *c-fgr* mRNA increased by at least 50-fold in response to EBV infection. Based on other data obtained with probes of similar specific activity¹³, we estimate that the EBV-infected lines contain <20 copies of *c-fgr* mRNA per cell.

The correlation of increased *fgr* proto-oncogene expression with EBV infection of transformed lymphocytes was nearly absolute. Increased concentrations of *c-fgr* mRNA were detected in Burkitt's lymphoma cells whether infected naturally or deliberately in tissue culture. In addition, each of the umbilical cord or peripheral blood lymphocyte cell lines established by EBV infection expressed *c-fgr* mRNA. Increases in the levels of transcripts related to other mammalian tyrosine kinase-coding *onc* genes (*v-fms* and *v-fes*) were also detected in some EBV-infected Burkitt's lymphoma cell lines (data not shown), but these RNAs were not consistently detected in EBV-infected lines and were, in some cases, detectable in uninfected cell lines. Taken together, these results establish a specific relationship between EBV infection and expression of the *fgr* proto-oncogene, which suggests that EBV infection results in transcriptional activation of *c-fgr*. Further studies will be required to show whether increased transcription rates alone are responsible for the increase in steady-state concentration of *c-fgr* mRNA.

Integration of EBV into host chromosomal DNA of Namalwa and IB4 cell lines has been described elsewhere¹⁴. In Namalwa cells, the site of integration is close to the *fgr* proto-oncogene locus on chromosome 1 (refs 7, 14), whereas in IB4 cells the

Table 1 Detection of *fgr* proto-oncogene mRNA in cell lines derived from Burkitt's and non-Burkitt's undifferentiated lymphomas

Histopathology	Cell line	EBNA status	Chromosome translocation	Geographical origin	Expression of <i>v-fgr</i> -related transcript*
Burkitt's lymphoma	Namalwa	+	8; 14	Africa	+++
	Keeper	+	8; 14	S. America	+++
	AG876	+	8; 14	Africa	++
	JI	+	2; 8	Europe	++
	Raji	+	8; 14	Africa	++
	LY47	+	8; 22	Africa	+
	BL2	—	8; 22	Europe	—
	ST486	—	8; 14	USA	—
	MC116	—	8; 14	USA	—
	CA46	—	8; 14	S. America	—
	EW36	—	8; 14	USA	—
	DS179	—	8; 14	USA	+
Non-Burkitt's lymphoma	JD38	—	8; 14	USA	—
	JLPC119	—	8; 14	USA	—

The characteristics of the cell lines examined have been described previously^{19,20}. The concentration of 3-kb *fgr*-related mRNA was determined by analysing 10 µg of poly(A)⁺ RNA in formamide-formaldehyde agarose gels as described²¹. The relative amounts of hybridizable RNA transferred to nitrocellulose filters were determined by rehybridization with a GR-FeSV-derived actin-specific probe.

* Detectable levels were graded by exposing hybridized nitrocellulose filters for various times as follows: +++, detected after overnight exposure; ++, detected after 2 days of exposure; +, detected after 5 days of exposure; —, not detectable after 5 days of exposure.

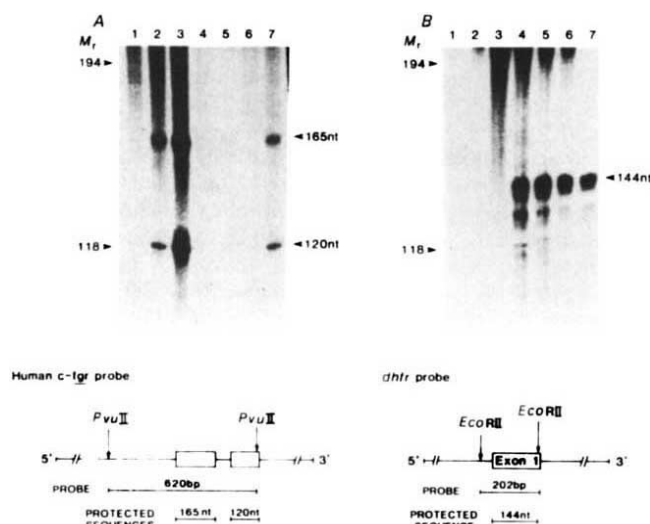


Fig. 3 Quantitation of *fgr* proto-oncogene mRNA induction in response to EBV infection. DNA fragments derived from the region 6.0–6.62 kb of the human *fgr* proto-oncogene⁷ (A) or the 5' end of the dihydrofolate reductase gene¹² (B), were cloned into M13 phage vectors and used as templates for the synthesis of uniformly labelled anti-sense DNA probes. Conditions for hybridization of total cellular RNA with the DNA probe, S_1 nuclease digestion and polyacrylamide gel electrophoretic analysis of protected sequences were as described elsewhere^{26,27}. Samples tested included 5 μ g (lanes 2) or 20 μ g (lanes 3) total cellular RNA from normal human peripheral blood mononuclear cells, 20 μ g RNA from uninfected Ramos (lanes 4) or BJAB (lanes 6) cells, as well as P3HR1-infected Ramos (lanes 5) or B95-8-infected BJAB (lanes 7) cell lines. Negative controls consisted of labelled probe hybridized in the absence of RNA and treated with S_1 nuclease (lanes 1). As expected, no *dhfr* mRNA was detected in peripheral blood mononuclear cells as these cells were not in cycle. The diagrams at the bottom of the figure show human *c-fgr* and *dhfr* exons and their protected sequences. The extent and locations of the human *c-fgr* exons shown were confirmed by nucleotide sequence analysis.

EBV genome integrates into chromosome 4. In the case of the Namalwa cell line, EBV may have a *cis*-acting effect on the expression of the *fgr* proto-oncogene. However, in most cases the influence of EBV on *c-fgr* is more likely to be *trans* as the EBV genome typically exists in the host as an episome. Thus, the isolation of EBV genes having specific functions may make it possible to identify the region of the EBV genome responsible for the increase in *fgr* proto-oncogene mRNA.

Several lines of evidence suggest that Burkitt's lymphoma results from a multi-step process involving the alteration of several genetic elements¹⁵. Epidemiological data suggest that EBV has an important role in the aetiology of African Burkitt's lymphoma¹⁶. Molecular studies have strongly implicated alterations of *myc* proto-oncogene expression in the process leading to American as well as African Burkitt's malignancies^{17,18}. However, there is no evidence to indicate that such genetic alterations are the only changes required to achieve lymphoid cell transformation. From the results of the present study we conclude that increased concentrations of *c-fgr* mRNA alone are not sufficient to induce malignant transformation of normal lymphocytes or even to immortalize them as continuous cell lines. Moreover, we have no direct evidence that EBV-induced immortalization involves *c-fgr*. Nevertheless, our results are consistent with the possibility that one step in EBV-induced B-cell immortalization involves transcriptional activation of the *fgr* proto-oncogene in response to an EBV-encoded function.

We thank N. Ellmore, L. Mickley and C. Lengel for technical assistance; N. Raab-Traub, I. T. Magrath and D. Ablashi for providing cell lines; T. Kawakami for making unpublished

nucleotide sequence data available to us; T. Shimada and A. Nienhuis for the *dhfr* probe; and S. Aaronson for helpful discussions.

Note added in proof: Recent experiments have revealed that *c-fgr* mRNA is not expressed in normal resting or activated B lymphocytes.

Received 8 July; accepted 24 October 1985.

- Downward, J. *et al.* *Nature* **307**, 521–527 (1984).
- Ulrich, A. *et al.* *Nature* **313**, 756–761 (1985).
- Rasheed, S., Barbacid, M., Aaronson, S. A. & Gardner, M. B. *Virology* **117**, 238–244 (1982).
- Naharro, G., Dunn, C. Y. & Robbins, K. C. *Virology* **125**, 502–507 (1983).
- Naharro, G. *et al.* *J. Virol.* **47**, 611–619 (1983).
- Naharro, G., Robbins, K. C. & Reddy, E. P. *Science* **223**, 63–66 (1984).
- Tronick, S. R. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **82**, 6595–6599 (1985).
- Henle, W. *et al.* *Science* **157**, 1064–1065 (1967).
- Pope, J. H., Horne, M. K. & Scott, W. *Int. J. Cancer* **3**, 857–866 (1968).
- Miller, G., Robinson, J., Heston, L. & Lipman, M. *Proc. natn. Acad. Sci. U.S.A.* **71**, 4006–4010 (1974).
- Klein, G. *et al.* *Int. J. Cancer* **18**, 639–652 (1976).
- Chen, M.-J. *et al.* *J. biol. Chem.* **259**, 3933–3943 (1984).
- Ley, T. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **81**, 6618–6622 (1984).
- Henderson, A., Ripley, S., Heller, M. & Kieff, E. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1987–1991 (1983).
- Klein, G. & Klein, E. *Nature* **315**, 190–195 (1985).
- Geser, A. *et al.* *Int. J. Cancer* **29**, 397–400 (1982).
- Dalla Favera, R. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **79**, 7824–7827 (1982).
- Taub, R. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **79**, 7837–7841 (1982).
- Lenoir, G. M., Preud'Homme, J. L., Bernheim, A. & Berger, R. *Nature* **298**, 474–476 (1982).
- Benjamin, D. *et al.* *J. Immun.* **129**, 1336–1342 (1982).
- Lehrach, H., Diamond, D., Wozney, J. M. & Bodtker, H. *Biochemistry* **16**, 4743–4751 (1977).
- Adams, S. L. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **74**, 3399–3403 (1977).
- Giard, D. J. *et al.* *J. natn. Cancer Inst.* **51**, 1417–1423 (1973).
- Clements, G. B., Klein, G. & Povey, S. *Int. J. Cancer* **16**, 125–133 (1975).
- Klein, G. *et al.* *Intervirology* **5**, 319–334 (1975).
- Berk, A. J. & Sharp, P. A. *Cell* **12**, 721–732 (1977).
- Ley, T. J., Anagnou, N. P., Pepe, G. & Niehuis, A. W. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4774–4779 (1982).

Expression of cauliflower mosaic virus reverse transcriptase in yeast

Hiroshi Takatsuji, Hirohiko Hirochika, Takao Fukushima* & Joh-E Ikeda

Laboratory of Molecular Genetics, National Institute of Agrobiological Resources, Tsukuba Science City, Yatabe, Ibaraki 305, Japan

* Institute of Applied Biochemistry, Tsukuba University, Tsukuba Science City, Ibaraki 305, Japan

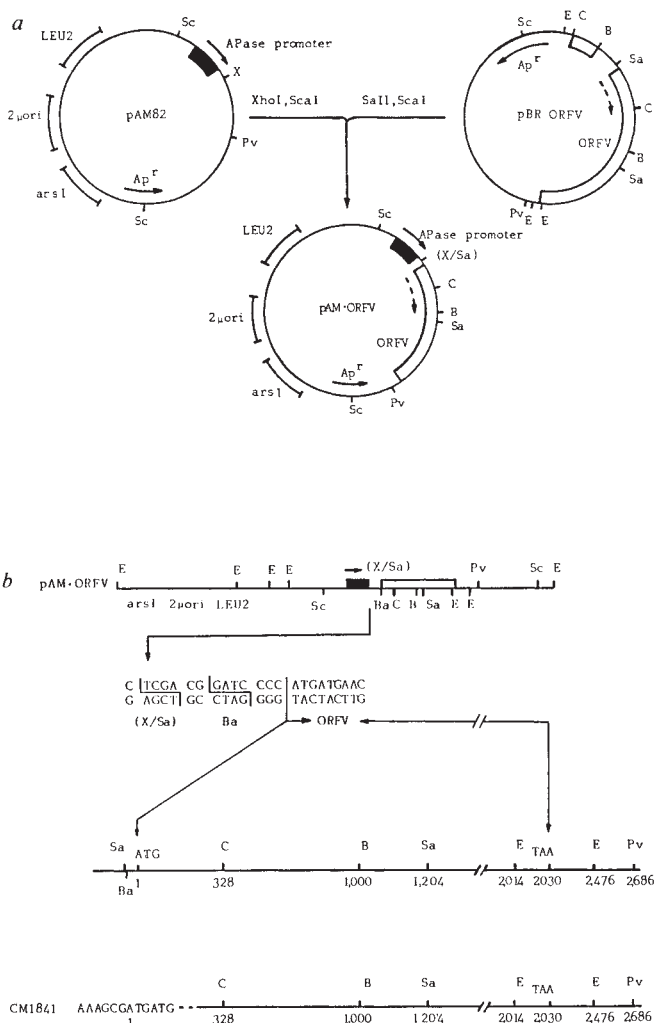
Cauliflower mosaic virus (CaMV) is a double-stranded DNA plant virus^{1–3} for which evidence has been presented that its replicative cycle involves a reverse transcription step^{4–14}. Until now, however, there has been no direct evidence for such a step. The sequence of the CaMV genome contains a number of open reading frames, one of which (ORF V) encodes a protein of predicted sequence homologous to retroviral reverse transcriptases¹⁰. We have cloned this gene, and expressed it in the yeast *Saccharomyces cerevisiae*. We report here that yeast expressing this gene accumulate significant levels of reverse transcriptase activity. This provides strong evidence that CaMV really does replicate through an RNA intermediate, converted to DNA by reverse transcriptase.

To obtain the direct expression of the ORF V gene in yeast cells, we constructed the plasmid pAM.ORFV in which the entire ORF V gene is inserted immediately downstream from the acid phosphatase (APase) promoter of the yeast expression vector pAM82 (ref. 15) (Fig. 1b). The resulting Leu⁺ yeast transformant AH22/pAM.ORFV was obtained. The integrity of the ORF V gene in pAM.ORFV was confirmed by fine restriction mapping of a rescued pAM.ORFV from AH22/pAM.ORFV cells (Fig. 1b).

AH22/pAM.ORFV and AH22/pAM82 cells were grown in a liquid medium and then induced in phosphate-deprived medium. The extract from AH22/pAM.ORFV cells, but not

Fig. 1 *a*, Construction of yeast expression plasmid pAM.ORFV. Restriction sites for the enzymes *Bam*HI (Ba), *Bgl*II (B), *Cla*I (C), *Eco*RI (E), *Pvu*II (Pv), *Sal*I (Sa), *Sca*I (Sc) and *Xho*I (X) are indicated. The fusion junction of *Xho*I and *Sal*I sites is shown as (X/Sa). Solid boxes represent the APase promoter region; open boxes represent the ORF V sequence. An arrow above the APase promoter shows the direction of transcription started from the APase promoter. A broken arrow shows the direction of translation started from the ATG initiator codon of the ORF V gene. Ap^r represents the ampicillin resistance gene. *b*, Fine restriction map of pAM.ORFV rescued from AH22/pAM.ORFV cells. All symbols represented are the same as indicated in *a*.

Methods. *a*, ORF V gene was recovered from pCaMV10 (ref. 29), which contains a full-length copy of CaMV strain CM1841 cloned at the *Sal*I site of pBR322. In the plasmid, the ORF V gene was contained in a form interrupted at the *Sal*I site by pBR322 sequences. The *Cla*I/*Sal*I fragment containing the 5' half of ORF V and the *Pvu*II/*Sal*I fragment containing the whole 3' half of ORF V were excised from pCaMV10, re-ligated at the *Sal*I site and cloned into the *Cla*I/*Pvu*II site of pBR322 (pBR.ORFV-D). The *Pst*I fragment containing the whole 5' half of ORF V was excised from pCaMV10 and 'nibbled' with *Bal*31 to the nucleotide next to the first ATG in two adjacent putative initiation codons, followed by *Bgl*II digestion. This fragment and the *Eco*RI/*Bgl*II fragment, which contains the 3' half of ORF V from pBR.ORFV-D, were inserted together into the *Sma*I/*Eco*RI site of M13mp8. pBR.ORFV, which contains the entire ORF V gene in correct orientation, was obtained by the insertion of the *Bgl*II fragment of the recombinant of mp8 and ORF V into the *Bgl*II site of pBR.ORFV-D. The *Sal*I/*Sca*I fragment of pBR.ORFV, containing the entire ORF V gene, was inserted into the *Xho*I/*Sca*I site of pAM82. The *Sal*I terminus of the former fragment was joined to the *Xho*I site of pAM82 at a position downstream from the APase promoter. The resultant plasmid, pAM.ORFV, was directly transformed into yeast AH22 cells. Transformation of yeast was performed by the modified method of Beggs³⁰ as described by Sherman *et al.*³¹. *Leu*⁺ transformants were selected on high-*P_i* medium containing 2% agar. *b*, *E. coli* C600 (*M*⁻, *R*⁻) was transformed with the plasmid DNA fraction which was prepared from spheroplasted AH22/pAM.ORFV cells (2×10^7 cells). The rescued pAM.ORFV DNAs were prepared from eight clones out of the ampicillin-resistant transformants (1×10^3 per μ g DNA) and subjected to fine restriction mapping.



from AH22/pAM82 cells, revealed an inducible RNA-dependent DNA polymerase activity which could use poly(rC)-oligo(dG) as well as poly(rA)-oligo(dT) as template-primer (Table 1). Eukaryotic DNA polymerases β and γ are known to use poly(rA)-oligo(dT), but not poly(rC)-oligo(dG)¹⁶. The low levels of activity observed with the extract from non-induced AH22/pAM82 cells could be explained by the presence of a β -type DNA polymerase observed previously in yeast¹⁷. The activity seen on poly(rC)-oligo(dG) with the extract from induced AH22/pAM.ORFV cells corresponds to the distinctive template specificity of a retroviral reverse transcriptase¹⁸. Negligible amounts of ³H-dGMP were incorporated using poly(rA)-oligo(dT) as a template, eliminating the possible presence or induction of terminal deoxynucleotidyl transferase. Our experiments also ruled out either induction or activation of a reverse transcriptase encoded by the host. An RNA-dependent DNA polymerase activity induced specifically in AH22/pAM.ORFV cells was therefore identified as the ORF V-encoded reverse transcriptase.

The specific activity of the ORF V protein produced in yeast by AH22/pAM.ORFV was purified about 4,000-fold over crude extracts, with a yield of 20% (Table 2a). During glycerol gradient centrifugation, the enzyme activity sedimented as a single peak of apparent relative molecular mass (*M_r*) ~70,000 (Fig. 2); this is consistent with the ORF V coding capacity of 78,000. Pfeiffer and Hohn¹⁹ have found that a 79,000-*M_r* polypeptide in organelles extracted from CaMV-infected turnips showed reverse transcriptase-like activity. The molecular size of our purified ORF V-encoded polymerase in its native condition is similar to that found in the virus-infected turnips. Therefore,

CaMV-encoded reverse transcriptase can be active as a monomeric polypeptide, as is murine leukaemia virus (MuLV)-encoded reverse transcriptase (reviewed in ref. 20). However, in contrast to the findings of Pfeiffer and Hohn¹⁹, the purified ORF V-encoded polymerase (glycerol gradient fraction) failed to show any activity on SDS-polyacrylamide gel electrophoresis. Furthermore, because of the very small yield of the purified ORF V-encoded polymerase, it is hard to estimate the size of the polymerase by SDS-polyacrylamide gel electrophoresis.

The purification of the CaMV reverse transcriptase was also monitored by examining the template specificity of the enzyme fraction at each purification step (Table 2b). Regardless of the step of purification, substantial activity was obtained by poly(rC)-oligo(dG) as well as with poly(rA)-oligo(dT). CaMV reverse transcriptase was also characterized as a MuLV-type reverse transcriptase by virtue of its activity on poly(rC)-oligo(dG) and its divalent metal ion requirements²¹. CaMV reverse transcriptase requires 2 mM MnCl₂ (instead of 5 mM MgCl₂) for maximum polymerization activity, and the reaction proceeds linearly for 2 h at an optimum temperature of 20 °C (data not shown). For reverse transcription with native RNA, for example, human liver messenger RNAs primed with oligo(dT) and tobacco mosaic virus RNA primed with synthetic DNA (25 bases), which is complementary to a 3' end of the viral RNA, Mn²⁺ was more effective than Mg²⁺ in promoting the synthesis of longer complementary DNA (in preparation), and also for promoting DNA-dependent DNA polymerase activity with native DNA such as single-stranded mp8 DNA primed with synthetic sequencing primer (in preparation). Neither endo- nor exodeoxyribonuclease activity other than

Table 1 RNA-dependent DNA polymerase activity in crude extracts of yeast AH22/pAM.ORFV and AH22/pAM82

Transformant	Induction	dNMP incorporation using different template-primers (pmol per mg protein)		
		poly(rA)-oligo(dT) ³ H-dTMP	poly(rC)-oligo(dG) ³ H-dGMP	poly(rA)-oligo(dT) ³ H-dGMP
AH22/pAM82	—	60.2	14.3	ND
AH22/pAM82	+	12.2	11.1	10.8
AH22/pAM.ORFV	—	34.6	14.2	ND
AH22/pAM.ORFV	+	486.0	191.0	13.0

The growth of yeast cells and induction were performed as described elsewhere¹⁵. Yeast cells were grown in high-inorganic phosphate (P_i) medium at 30 °C with aeration. When the cell density reached $\sim 1-2 \times 10^6$ cells ml⁻¹, the cells were centrifuged and resuspended in low- P_i medium for induction. Control samples were treated similarly, with the exception that high- P_i medium was used. When the density reached $2-3 \times 10^7$ cells ml⁻¹, 50-ml samples of induced (+) and non-induced (—) cultures were centrifuged. The cells were suspended in a 15-ml solution of Zymolyase (200 μ g ml⁻¹), 1.2 M sorbitol, 50 mM Tris-HCl (pH 7.5), 14 mM 2-mercaptoethanol and incubated for 30 min at 30 °C. The spheroplasts were collected by centrifugation and lysed in 2 ml of 50 mM Tris-HCl (pH 7.5), 0.1% Triton X-100, 1 mM dithiothreitol. The lysate was clarified by centrifugation at 10,000g for 10 min. Protein concentration was measured by the modified Bradford method²⁶ using a Bio-Rad protein assay kit. RNA-dependent DNA polymerase activity was assayed as described by Marcus *et al.*²⁷ and Dion *et al.*²⁸ with modifications. Unless otherwise stated, the reaction mixture (0.05 ml) contained 50 mM Tris-HCl (pH 7.5), 50 mM KCl, 2 mM MnCl₂, 80 μ g ml⁻¹ poly(rA), 10 μ g ml⁻¹ oligo(dT), 4 μ M ³H-dTTP (3,000–4,000 c.p.m. pmol⁻¹), 5 mM dithiothreitol, 0.01% Nonidet P-40 (NP40), 0.1 mg ml⁻¹ bovine serum albumin and 9 μ g protein of the crude extracts, and was incubated for 30 min at 30 °C. When indicated, poly(rC)-oligo(dG) (90 μ g ml⁻¹) or poly(rCm)-oligo(dG) (90 μ g ml⁻¹) was substituted for poly(rA)-oligo(dT) and ³H-dGTP (2,000 c.p.m. pmol⁻¹) was substituted for ³H-dTTP. Reactions were halted by the addition of 0.1 ml of herring sperm DNA (1 mg ml⁻¹), 0.1 ml of 0.1 M sodium pyrophosphate and 3 ml of 10% (w/v) trichloroacetic acid and kept on ice for 10 min. Radioactivities incorporated into acid-insoluble fractions were counted. ND, not done.

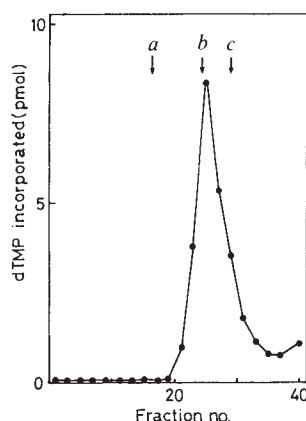


Fig. 2 Glycerol gradient centrifugation of CaMV reverse transcriptase. The enzyme fraction of step III was sedimented through a glycerol gradient as described in Table 2 legend. Aliquots (0.12 ml) of the fractions were collected from the bottom of the tube and 2- μ l aliquots assayed for RNA-dependent DNA polymerase activity using poly(rA)-oligo(dT) as template-primer. The positions of the marker proteins phosphorylase *b* (M_r 92,500) (*a*), bovine serum albumin (68,000) (*b*) and ovalbumin (43,000) (*c*) in a parallel gradient are indicated.

ribonuclease activities, which seriously affected cDNA synthesis with native RNAs, was associated with reverse transcriptase in a glycerol gradient fraction. This finding might be expected in view of the absence of a putative endonuclease coding domain in CaMV ORF V²².

We have also attempted to express the CaMV ORF V gene in *Escherichia coli*. The entire ORF V gene, which lacks the first of two adjacent putative initiation codons, was inserted immediately downstream from the initiation codon of the *tac* promoter in the *E. coli* expression vector ptac12. The recombinant plasmid was transformed into *E. coli* NK5830 cells and the resulting

Table 2 Purification of ORF V gene product induced in yeast cells (*a*) and template specificity of enzyme fraction in each purification step (*b*)

<i>a</i>	Fractionation procedure	Total protein (mg)	Total units	Specific activity
I	Crude extract	90	105.3	1.18
II	Phosphocellulose	2.52	37.1	15.1
III	Denatured-DNA cellulose	0.41	23.4	74.9
IV	Glycerol gradient	<0.005	20.2	>4,032.0

<i>b</i>	Fractionation procedure	Specific activity		
		poly(rA)-oligo(dT)	poly(rC)-oligo(dG)	poly(rCm)-oligo(dG)
I	Crude extract	1.21	0.418	ND
II	Phosphocellulose	20.1	2.82	1.41
III	Denatured-DNA cellulose	87.7	13.7	9.23
IV	Glycerol gradient	4,050.1	1,030.0	1,050.3

a, CaMV reverse transcriptase was purified from 50 ml of crude extract (step I, 90 mg total protein). The crude extract was prepared from the spheroplasts obtained from 1.5 l culture of induced yeast AH22/pAM.ORFV cells, as described in Table 1 legend. Step I product was applied to a phosphocellulose column (2 cm $\times$ 3 cm) equilibrated with buffer A (50 mM Tris-HCl (pH 7.5), 2 mM 2-mercaptoethanol, 0.2 mM EDTA, 0.01% NP40, 20% (v/v) glycerol) containing 50 mM KCl. The column was washed with 30 ml of the same buffer and the enzyme was eluted with a 100-ml linear gradient of 0.05–0.8 M KCl in buffer A. RNA-dependent DNA polymerase fractions (0.4–0.55 M KCl) were pooled and dialysed against buffer A containing 20 mM KCl (step II). Step II product (2.5 mg protein) was applied to a denatured DNA cellulose column (1.2 $\times$ 2.5 cm) equilibrated with buffer A containing 20 mM KCl. The column was washed with 5 ml of the same buffer and the enzyme was eluted with a 30-ml linear gradient of 0.02–1.0 M KCl in buffer A. The enzyme fractions (0.26–0.38 M KCl) were pooled (step III). Step III product (0.41 mg protein) was diluted with two volumes of buffer A and applied to a small phosphocellulose column (1.2 $\times$ 1 cm) equilibrated with buffer A. The concentration of the enzyme was carried out by stepwise elution with buffer B (50 mM Tris-HCl (pH 7.5), 2 mM 2-mercaptoethanol, 0.01% NP40, 0.2 mM EDTA, 10% (v/v) glycerol, 1.0 M KCl). The concentrated enzyme fraction was applied to a 5-ml glycerol gradient (15–30%) in buffer B. Centrifugation was for 20 h at 48,000 r.p.m. in an RPS-50-2 Hitachi rotor at 4 °C. Active fractions were pooled and dialysed against buffer A containing 50% (v/v) glycerol and 1.0 M KCl, and stored at –20 °C (step IV, 0.005 mg protein). After this step, no detectable loss of activity occurred over 3 months at –20 °C. Aliquots of 5 ng–9 μ g protein from each step of enzyme fraction were used for the RNA-dependent DNA polymerase assay with poly(rA)-oligo(dT), as described in Table 1 legend. One unit of the enzyme catalysed the incorporation of 1 nmol of dTMP in 60 min at 20 °C. *b*, Each step of enzyme fraction (5 ng–9 μ g protein) was assayed for RNA-dependent DNA polymerase using poly(rA)-oligo(dT), poly(rC)-oligo(dG) and poly(rCm)-oligo(dG) as described in Table 1 legend. The incubation time at 20 °C was the same as above except that step I was for 15 min.

transformant, NK5830/ptac.ORFV, was treated with 1 mM isopropyl β -D-thiogalactoside to induce the ORF V product under control of the *tac* promoter. The induction-specific polypeptide with an apparent M_r of 70,000 was found in SDS-polyacrylamide gels (data not shown). The polypeptide was accumulated as 1–3% of the total protein in *E. coli* during an induction of 2 h. The M_r of the polypeptide induced in *E. coli* was nearly the same as that of a native CaMV reverse transcriptase produced in yeast cells. However, in *E. coli*, no reverse transcriptase-like activity was detected, even in the partially purified fractions.

Some other foreign proteins, such as the human hepatitis B surface antigen gene^{15,23}, have been difficult to express in active form in *E. coli*. It is conceivable that post-translational modification and/or slight processing which is not sufficient to affect the apparent size of the polymerase could be required to produce CaMV reverse transcriptase in active form. For example, a putative protease-coding domain has also been predicted in the CaMV ORF V gene by Toh *et al.*²². This putative protease may be required to process a reverse transcriptase to active form in yeast cells. Recently, MuLV reverse transcriptase, of which the NH₂ and COOH termini approximate the correct termini of the authentic protein, has been expressed in active form in *E. coli* cells^{24,25}. In the case of our cloning of CaMV reverse transcriptase, we did not leave any coding sequences

between the APase or *tac* promoter and the initiation codon of the ORF V gene in the expression plasmids, and so the products should be a non-fused, complete ORF V polypeptide. This has yet to be confirmed by amino-acid sequence studies.

We thank Dr R. J. Shepherd for pCaMV10, Dr K. Matsubara for the yeast expression vector pAM82 and yeast AH22, Mr T. Noguchi for his help in restriction mapping, and Dr T. R. Broker for his reading of the manuscript. This work is supported in part by Funds for Promoting Science and Technology from the Science and Technology Agency of Japan, and Grants-in-Aid for Cancer Research and for Cooperative Research from the Ministry of Education, Science and Culture of Japan.

Received 8 August; accepted 30 October 1985.

1. Shepherd, R. J. *A. Rev. Pl. Physiol.* **30**, 405-423 (1974).
2. Hohn, T., Richards, K. & Lebeurier, G. *Curr. Topics Microbiol. Immun.* **96**, 194-236 (1982).
3. Howell, S. H. *A. Rev. Pl. Physiol.* **33**, 609-650 (1982).
4. Varmus, H. E. *Nature* **304**, 116-117 (1983).
5. Pfeiffer, P. & Hohn, T. *Cell* **33**, 781-789 (1983).
6. Hull, R. & Covey, S. N. *Trends biochem. Sci.* **8**, 119-123 (1983).
7. Thomas, C. M., Hull, R., Bryant, J. A. & Maule, A. J. *Nucleic Acids Res.* **13**, 4557-4576 (1985).
8. Boeke, J. D., Garfinkel, D. J., Styles, C. A. & Fink, G. R. *Cell* **40**, 491-500 (1985).
9. Shibata, T. & Saigo, K. *Nature* **302**, 119-124 (1983).
10. Toh, H., Hayashida, H. & Miyata, T. *Nature* **305**, 827-829 (1983).
11. Flavell, A. J. *Nature* **310**, 514-516 (1984).
12. Saigo, K. *et al. Nature* **312**, 659-661 (1984).
13. Summers, J. & Mason, W. S. *Cell* **29**, 403-415 (1982).
14. Hirochika, H., Takatsuiji, H., Ubasawa, A. & Ikeda, J.-E. *EMBO J.* **4**, 1673-1680 (1985).
15. Miyano, H. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 1-5 (1983).
16. Gerald, G. F., Rottman, F. & Green, M. *Biochemistry* **13**, 1632-1641 (1974).
17. Chang, L. S. *J. biol. Chem.* **252**, 1873-1880 (1977).
18. Verma, I. M. *Biochim. biophys. Acta* **473**, 1-38 (1977).
19. Hohn, T., Hohn, B. & Pfeiffer, P. *Trends biochem. Sci.* **10**, 205-209 (1985).
20. Dickson, C., Eisenman, R., Fan, H., Hunter, E. & Teich, N. in *Molecular Biology of Tumor Viruses* (eds Weiss, R. A., Teich, N., Varmus, H. E. & Coffin, J. M.) 513-648 (Cold Spring Harbor Laboratory, New York, 1982).
21. Kohno, M. & Ishihama, A. *Eur. J. Biochem.* **97**, 257-266 (1978).
22. Toh, H. *et al. EMBO J.* **4**, 1267-1272 (1985).
23. Valenzuela, P., Mediana, A., Rutter, W. J., Ammerer, G. & Hall, B. D. *Nature* **298**, 347-350 (1982).
24. Tanase, N., Roth, M. & Goff, S. P. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4944-4948 (1985).
25. Kotewicz, M. L., D'Alessio, J. M., Driftmier, K. M., Blodgett, K. P. & Gerard, G. F. *Gene* **35**, 249-258 (1985).
26. Bradford, M. M. *Analyt. Biochem.* **72**, 248-254 (1976).
27. Marcus, S. L., Modak, M. J. & Cavalieli, L. F. *J. Virol.* **14**, 853-859 (1974).
28. Dion, A. S., Vaidya, A. B., Fout, G. S. & Moore, D. H. *J. Virol.* **14**, 40-46 (1974).
29. Gardner, R. C. *et al. Nucleic Acids Res.* **9**, 2871-2888 (1981).
30. Beggs, J. D. *Nature* **275**, 104-109 (1978).
31. Sherman, F., Fink, G. R. & Hicks, J. B. *Methods in Yeast Genetics*, 112-115 (Cold Spring Harbor Laboratory, New York, 1982).

Detection of cytosine methylation in the maize alcohol dehydrogenase gene by genomic sequencing

Harry Nick*, Barbara Bowen*, Robert J. Ferl† & Walter Gilbert*

Biogen Research Corporation, Cambridge, Massachusetts 02142, USA

† Department of Botany, University of Florida, Gainesville, Florida 32611, USA

DNA methylation is believed to be involved in the control of gene expression^{1,2}. In higher plants, up to 30% of the cytosine residues can be methylated—at C-X-G trinucleotides as well as C-G dinucleotides—compared with only 2-8% of total cytosines appearing exclusively as 5-methyl-C-5-methylcytosine in higher animals¹⁻⁴. We have used the genomic sequencing technique of Church and Gilbert⁵ to study the methylation of an alcohol dehydrogenase gene, *Adhl*, from maize. *Adhl* is one of two unlinked maize alcohol dehydrogenase genes^{6,7} and appears in several tissues, including the scutellum of the kernel and the primary root of the seedling. It is further induced in primary roots subjected to anaerobiosis⁸⁻¹⁰.

However, this gene is totally repressed in maize leaves¹¹. Nonetheless, we have now found that a 900-base-pair (bp) region 5' to the ATG is not methylated in leaves, even though it is rich in potential methylation sites. The first methylations occur 965 bp 5' to the ATG. Thus, methylation does not appear to be involved in repressing this plant gene.

Restriction endonucleases whose activity is blocked when their recognition site is methylated have been used to detect modified C-G sites in the genome^{12,13}. However, even with the ability to use additional enzymes such as *PstI* and *PvuII* (ref. 14) for methylation analysis in plants, restriction sites represent a small percentage of potential methylation sites. To demonstrate the applicability of genomic sequencing to the maize genome, which, at $\sim 7 \times 10^9$ bp, is two to three times larger than the mammalian genome^{15,16}, we analysed the methylation levels of cytosine residues in the 5'-flanking region of *Adhl*.

Figure 1 illustrates the genomic sequencing strategy. After cutting genomic DNA with a restriction enzyme that acts near a region of interest, the DNA was reacted with hydrazine and piperidine to liberate fragments terminating at cytosine

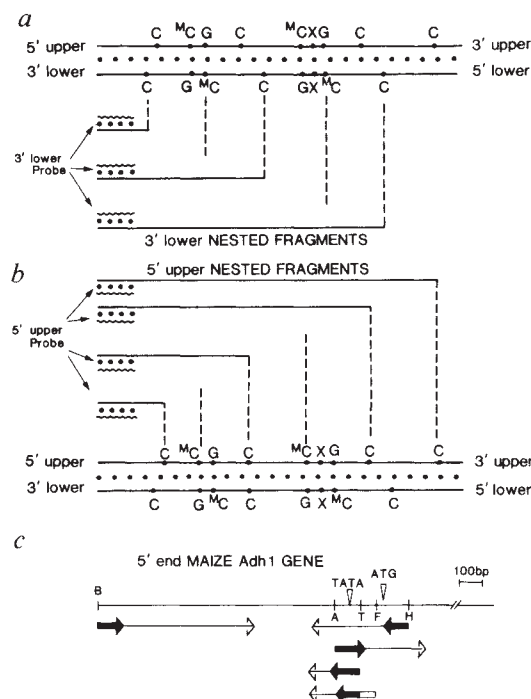


Fig. 1 *a*, A double-stranded genomic restriction fragment showing upper and lower strands in a 5' to 3' convention. A collection of nested fragments are generated by a partial cleavage with hydrazine and piperidine specific for cytosine residues. Methylation of cytosine prevents reaction with hydrazine so that no fragment is generated which terminates at this residue and thus the corresponding band on a sequencing gel is absent. Four possible strand-specific end-labelled sequences may be displayed. A 3' lower probe illustrates the detection of the 3' lower nested fragments. *b*, The detection of the opposite strand sequence originating from the same genomic restriction fragment. The hybridization with a 5' upper probe displays the collection of 5' upper nested fragments terminating in cytosine residues. Similarly, sequence originating from either the 3' upper or 5' lower strands may be displayed with the corresponding 3' upper and 5' lower probes. *c*, The region at the 5' end of the maize *Adhl* gene probed for cytosine methylation using genomic sequencing. The ATG codon and TATA box are shown by open arrows. The five vertical lines illustrate the position of the genomic restriction sites (*Bam*HI (B), *Hae*III (A), *Taq*I (T), *Hin*FI (F) and *Hin*DI (H)), from which genomic sequence (thin lined arrow) was derived using the corresponding 100-bp single-strand DNA probes (thick darkened arrows). The open portion of one arrow shows how a probe starting from the *Taq*I site was used to probe *Hin*FI-cut chemically treated genomic fragments.

* Present addresses: Department of Biochemistry and Molecular Biology, College of Medicine, University of Florida, Gainesville, Florida 32610, USA (H.N.); Biological Laboratories, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA (B.B., W.G.).

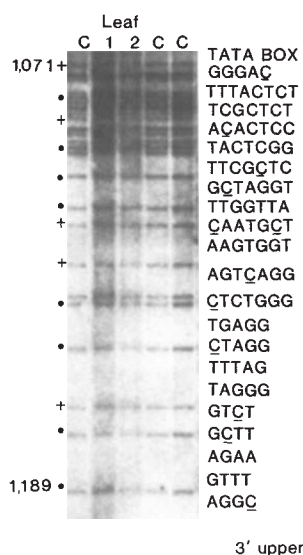


Fig. 2 Genomic sequence originating from the *Hind*III site extending towards the TATA box. Lane 1, 50 μ g leaf genomic DNA; lane 2, 25 μ g DNA. The lanes labelled C are hydrazine reactions on cloned plasmid DNA restricted with *Hind*III to illustrate the position of each cytosine residue. The sequence begins at position 1,189 (numbering scheme in Fig. 4) and proceeds 5' to the TATA box near position 1,071. The dots and crosses denote the positions of cytosines in C-G dinucleotides and C-X-G trinucleotides, respectively.

Methods. DNA was isolated from maize leaves by a modification of published procedures²⁷; the nucleus isolation buffer was 0.3 M sucrose, 50 mM Tris-HCl pH 8.0, 5 mM MgCl₂, 15 mM NaCl, 0.1 mM phenylmethylsulphonyl fluoride, 0.1 mM EGTA and 0.5 mM 2-mercaptoethanol. The crude nuclear pellet was lysed in 50 mM Tris-HCl pH 8.0, 20 mM EDTA, 1% sarkosyl and the DNA purified by equilibrium density centrifugation in CsCl. Leaf genomic DNA was digested to completion with *Hind*III, reacted with hydrazine and piperidine¹⁷, applied to a denaturing gel, electrotransferred and the nylon ultraviolet irradiated as described elsewhere⁵. The 3' upper strand sequence is displayed by hybridization with a 3' upper single-strand DNA probe⁵. We have noted that identical single-copy maize signals may be obtained, with a corresponding lower noise level, when 0.25 mCi of label, rather than 1 mCi, is used. The hybridizations were performed at 65 °C as described above except that some of the filters were hybridized in a lexan tube submerged in water at 65 °C in a new rotating device, which will be described elsewhere (R. Tizard and H.N., manuscript in preparation). Washes were performed as described previously²⁸. The membranes were autoradiographed with an intensifying screen for 2.5 days to 2 weeks.

residues¹⁷. The fragments were then separated on a denaturing gel, electrotransferred to a non-charged nylon membrane and photo-crosslinked to the nylon. Hybridization with an end-specific, strand-specific DNA probe displayed the sequence of the desired genomic restriction fragment (the 5' upper or 3' lower probe in Fig. 1⁵). Because 5-methylcytosine does not react with hydrazine, no fragment was generated from this site¹⁸. In plants, this will result from methylation of either the sequence C-X-G or C-G (refs 3, 4 and Fig. 1a, b).

Figure 1c summarizes the probes used and the region studied. Hybridization with a 3' upper probe originating from a *Hind*III site in the first intron produced the sequence of the 3' upper strand from a *Hind*III digest of leaf genomic DNA. Figure 2 displays sequence from 25 and 50 μ g of leaf DNA, flanked by control cytosine reactions from cloned plasmid DNA. The sequence runs out 100 bases to the TATA box and covers 13 potential methylation sites. All of the bands corresponding to C-G to C-X-G sites are present in the genomic DNA, which demonstrates that the 5' end of *Adhl* is hypomethylated in leaves. Further data (Fig. 4) show that at least 350 bp in the 5' region

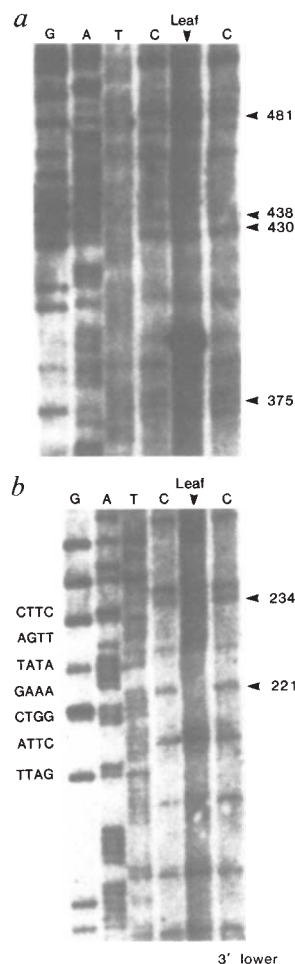


Fig. 3 Genomic sequence originating from the *Bam*HI site 1,200 bp 5' to the ATG codon. Parts a and b are from the same autoradiogram. Leaf genomic DNA (30 μ g) was digested to completion with *Bam*HI and treated as in Fig. 2. A 3' lower probe was hybridized with this membrane to display 3' lower strand sequence, which begins at position 192 and extends 3' towards position 481 at the top of a. a, The positions of four potential methylation sites (481, 438, 430 and 375) are shown by the arrows. All of these bands are detectable, as compared with the G, A, T and C reactions on the cloned plasmid DNA. b, The next two potential methylation sites (234 and 221) as indicated by the arrows are absent due to cytosine methylation at a C-A-G and C-T-G trinucleotide.

are hypomethylated in leaves.

We searched for methylation in an area further 5' to the gene by probing from a *Bam*HI site 1,200 bp from the gene's ATG codon. Figure 3a depicts the 3' lower strand sequence and the four potential methylation sites; these sites are also hypomethylated. However, Fig. 3b from the same autoradiogram reveals two methylated sites 950 bp upstream of the start site, at positions 234 and 221 in the trinucleotide sequences C-T-G and C-A-G. This experiment thus defines a transition in the state of methylation at the 5' end of the maize *Adhl* gene.

Figure 4 summarizes our data and gives the number of potential methylation sites per 100 bp. We have probed 61 out of 107 of the C-G and C-X-G sites within 1,400 bp surrounding the 5' end of *Adhl*. For 950 bp flanking the ATG codon, the *Adhl* gene is completely hypomethylated in leaves (with a single gap in our data at position 650-900; Fig. 4). We were unable to find probes that would yield clean sequence across the 650-900-bp region. We attribute this either to the fact that this region and its immediate vicinity are repetitious or that there is a universal nicking of the DNA strands in this region in the nuclear prepar-

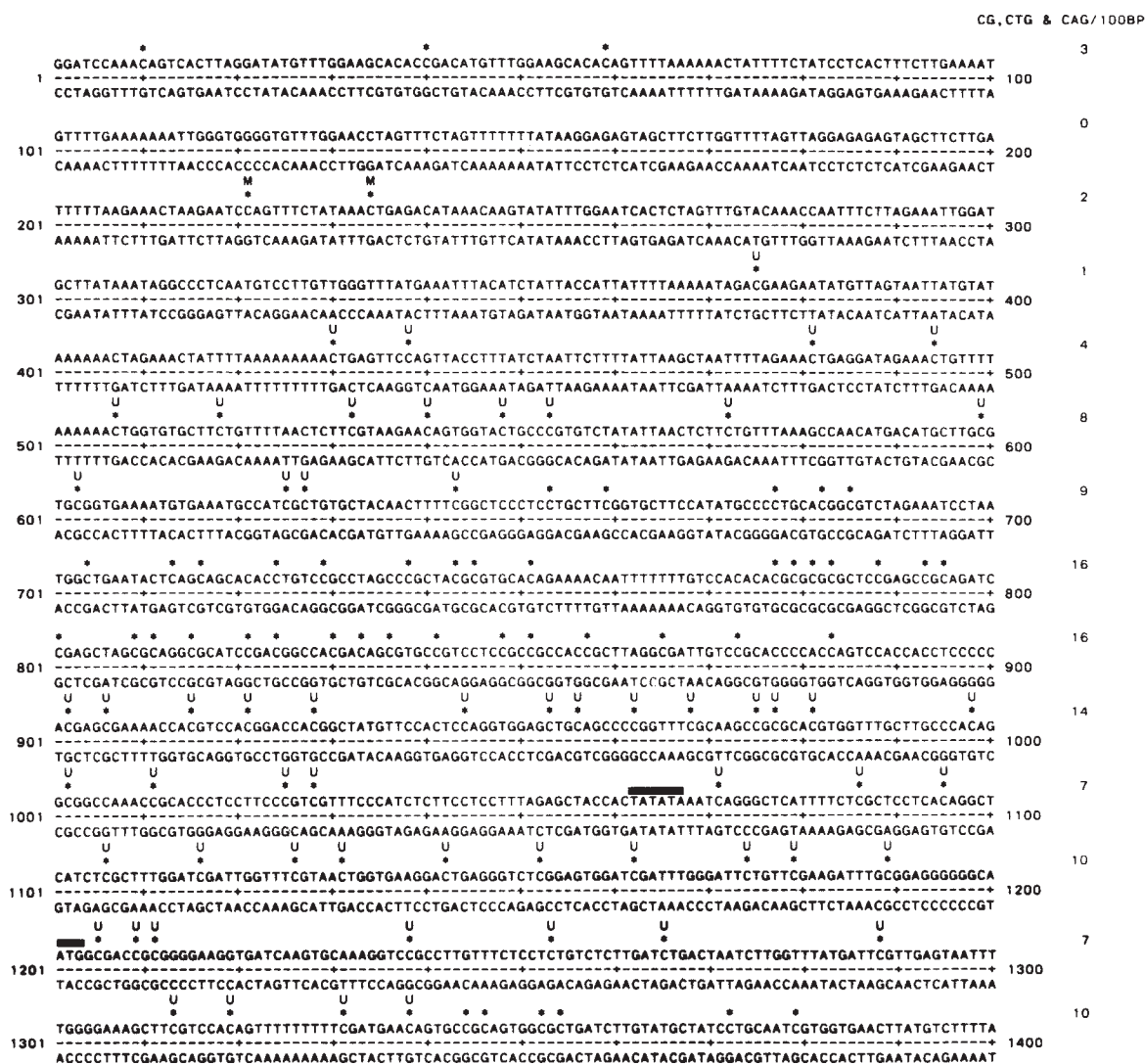


Fig. 4 Summary of cytosine methylation at the 5' end of the maize *Adhl* gene. An asterisk appears above each potential methylation site in the 1,400 bp of 5'-flanking sequence. Only three types of sites, C-G, C-A-G and C-T-G, are indicated. The TATA box is centred at position 1,065 and the ATG codon at position 1,201, each denoted by a dark line. The U designates sites which are unmethylated and the M sites which have been found to be methylated. A gap exists in our data between sites 650 and 900. Only a small amount of the primary data have been shown; all of the data are illustrated on the top strand for simplicity. The numbers in the column on the right give the number of C-G, C-A-G and C-T-G sites per 100 bp.

ations. We expect this 'gap' to be hypomethylated, since the flanking sequences are uniformly hypomethylated. At 130 bases further 5', we detect the first cytosine methylation; positions 234 and 221 are completely methylated.

The frequency of potential methylation sites in the first 300 bp of Fig. 4, the methylated region, is 1.3 sites per 100 bp. This is very low compared with the average of 9.3 sites per 100 bp in the next 1,100 bases, which is the statistical expectation and suggests that this last region is never methylated in any tissue and therefore never subjected to the pressures of transition mutagenesis through spontaneous deamination¹⁹⁻²¹.

Considerable data have been accumulated using modification-sensitive restriction endonucleases, predominantly with animal genes^{1,2}. The major implication of such data is that actively transcribed genes are hypomethylated relative to inactive genes—specifically, C-G sites at the 5' end of genes^{22,23}. The application of methylation-sensitive restriction enzymes in plants has previously been limited to studies of the T-DNA of crown gall tumours, repetitive sequences and 5S RNA genes^{14,24-26}. Our results are the first extensive methylation study in plants correlated to a unique gene sequence, *Adhl*, as well as

the application of genomic sequencing to a genome as large as that of maize.

Because the 950 bases in front of the gene, in a tissue in which this gene is completely repressed, are hypomethylated, we do not see any role for methylation in the control of expression of *Adhl*. If methylation were to function in repression here, it would do so from a distance of 1,000 bp.

We thank Marty Sachs and W. J. Peacock for sequence data before publication, M. Freeling for the genomic clone and Anna-Lisa Paul for technical assistance. This research was supported by USDA/CARGO grant 83-CRCR-1-1271 and Biogen Research Corporation.

Received 31 July; accepted 11 October 1985.

- Cooper, D. N. *Hum. Genet.* **64**, 315-333 (1983).
- Doefler, W. A. *Rev. Biochem.* **52**, 93-124 (1983).
- Gruenbaum, Y., Naveh-Man, T., Cedar, H. & Razin, A. *Nature* **292**, 860-862 (1981).
- Vanyushin, B. F. *Curr. Topics Microbiol. Immun.* **108**, 99-114 (1984).
- Church, G. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1991-1995 (1984).
- Schwartz, D. *Proc. natn. Acad. Sci. U.S.A.* **56**, 1431-1436 (1966).
- Dennis, E. S. *et al. Nucleic Acids Res.* **12**, 3983-4000 (1984).
- Ferl, R. J., Dlouhy, S. R. & Schwartz, D. *Molec. gen. Genet.* **169**, 7-12 (1979).
- Freeling, M. *Molec. gen. Genet.* **127**, 215-227 (1973).

10. Sachs, M. M., Freeling, M. & Okimoto, R. *Cell* **20**, 761-767 (1980).
11. Okimoto, R., Sachs, M. M. & Freeling, M. *Planta* **150**, 89-94 (1980).
12. Bird, A. & Southern, E. M. *J. molec. Biol.* **118**, 27-47 (1978).
13. Waalwijk, C. & Flavell, R. A. *Nucleic Acids Res.* **5**, 4631-4641 (1978).
14. Hepburn, A. G., Clarke, L. E., Pearson, L. & White, J. *J. molec. appl. Genet.* **2**, 315-329 (1983).
15. Fasman, G. D. (ed.) *Handbook of Biochemistry and Molecular Biology* 3rd edn, Vol. 2, 284-306 (1983).
16. Bennett, M. D. & Smith, J. B. *Phil. Trans. R. Soc. B274*, 227-274 (1976).
17. Maxam, A. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).
18. Ohmori, H., Tomizawa, J. & Maxam, A. M. *Nucleic Acids Res.* **5**, 1479-1486 (1978).
19. Coulondre, C., Miller, J. H., Farabaugh, P. J. & Gilbert, W. *Nature* **274**, 775-780 (1978).
20. Bird, A. *Nucleic Acids Res.* **8**, 1499-1504 (1980).
21. Barker, D., Schafer, M. & White, R. *Cell* **36**, 131-138 (1984).
22. Busslinger, M., Hurst, J. & Flavell, R. A. *Cell* **34**, 197-206 (1983).
23. Ott, M. *et al. Cell* **30**, 825-833 (1982).
24. Gelvin, S. B., Karcher, S. J. & DiRita, V. J. *Nucleic Acids Res.* **11**, 159-174 (1983).
25. Deumling, B. *Proc. natn. Acad. Sci. U.S.A.* **78**, 338-342 (1981).
26. Goldsborough, P. B., Ellis, T. H. N. & Lomonosoff, G. P. *Nucleic Acids Res.* **10**, 4501-4514 (1982).
27. Rivin, C., Zimmer, E. & Walbot, V. *Maize for Biological Research* (ed. Sheridan, W.) (University of North Dakota Press, Grand Forks, 1982).
28. Ephrussi, A., Church, G. M., Tonegawa, S. & Gilbert, W. *Science* **227**, 134-146 (1985).

Transcription factor Sp1 recognizes a DNA sequence in the mouse dihydrofolate reductase promoter

William S. Dynan*[‡], Shelley Sazer[†], Robert Tjian*
& Robert T. Schimke[†]

* Department of Biochemistry, University of California, Berkeley, California 94720, USA

[†] Department of Biological Sciences, Stanford University, Stanford, California 94305, USA

Human (HeLa) cells contain a host-cell-encoded transcription factor, Sp1, which is required for transcription of simian virus 40 (SV40) promoters^{1,2}. Since the discovery of Sp1 we have been interested in learning what role this factor plays in uninfected cells. A monkey cellular gene promoter interacts with Sp1³, but no gene products linked to this promoter have yet been identified. The finding that the sequence of the 5'-flanking DNA⁴ of the mouse dihydrofolate reductase (DHFR) gene contains several regions showing strong homology to the Sp1 binding region of simian virus 40 (SV40) prompted us to undertake experiments with *dhfr*. We report here that Sp1 binds to these regions in the *dhfr* promoter, and that Sp1-containing preparations stimulate transcription from the *dhfr* promoter in an *in vitro* reaction. Our results suggest that in addition to its interactions with the SV40 viral promoter, one function of Sp1 is to direct the expression of the cellular DHFR gene.

DHFR catalyses the production of tetrahydrofolate, which is required for *de novo* synthesis of thymidylate, purines and glycine. The level of *dhfr* RNA varies during the cell cycle, being maximal at the onset of DNA synthesis⁵. Recent studies indicate that the 5' end of the major *dhfr* messenger RNA lies 55 base pairs (bp) 5' of the translation start²⁵ while the previously described -115 terminus^{4,5} is a minor species in exponentially growing cells. As with SV40⁶ and the monkey cellular gene promoter⁷, transcription from the *dhfr* promoter region is bidirectional, with a transcript of unknown function synthesized from the opposite strand⁸.

Deletion analysis of the *dhfr* promoter has been used to localize sequences showing promoter activity. DNA sequences within 300 bp of the *dhfr* initiation codon direct sufficient expression for cell survival in selective medium⁴. When partially purified Sp1 from human (HeLa) cell extracts was incubated with an end-labelled *dhfr* fragment containing this promoter region, four discrete regions were protected from DNase I in a footprinting assay (Fig. 1). Each footprint covered 20 bp on each strand of the DNA and included one 'GC box', a decanucleotide

homologous to the consensus Sp1 binding sequence (Fig. 1a, b). To show that Sp1 was specifically responsible for the observed protection, the binding assay was carried out in the presence of several competitor DNAs (Fig. 2). The unlabelled *dhfr* fragment competed for protein bound to the *dhfr* promoter region, and almost eliminated the protection (Fig. 2a). A fragment containing the Sp1 binding region of the SV40 gene also reduced protection to a similar extent (Fig. 2b), suggesting that the same protein binds to the promoter regions of the SV40 gene and *dhfr*. By contrast, a fragment of similar size containing the adenovirus-2 (Ad-2) major late promoter, which has no Sp1 binding site, had no effect on the pattern of protection (Fig. 2c). The reciprocal experiment, where the various fragments were added to a binding reaction containing labelled SV40 probe, gave similar results (not shown).

In vitro transcription of *dhfr* directed by a HeLa nuclear extract was mapped in a runoff assay (Fig. 3a). Using two different templates, we saw transcripts (Fig. 3a, lanes 1, 3, arrows) of roughly the size expected for RNA initiating at -55, relative to the translational start codon. The position of the 5' end, 40 bp downstream from the most proximal Sp1 site, corresponds to the pattern seen in SV40 and herpesvirus transcription units (refs 2, 9, 10 and K. Jones, K. Yamamoto and R. T., unpublished; for review, see ref. 11). α -Amanitin (1 μ g ml⁻¹) inhibited synthesis, as expected for products of RNA polymerase II. When the *in vitro* RNA was analysed by primer extension (Fig. 3b, lane 1), one major group of bands was again seen, corresponding to initiation on the *dhfr* coding strand near -55 (Fig. 3b, large arrow); a less intense band was also visible, corresponding to initiation at -115 (small arrow). Several bands of smaller size, present in variable amounts, could represent either RNA initiating downstream of -55, or incomplete complementary DNA products synthesized from the major *dhfr* transcript. A similar primer extension analysis carried out using *in vivo* RNA (Fig. 3b, lane 2) produced essentially identical results, showing that the *in vitro* transcription system faithfully reproduced the *in vivo* pattern of RNA synthesis.

These experiments were then extended to a more highly purified transcription system, where Sp1 had been separated from the other enzymatic components required for transcription initiation. Without Sp1, *dhfr*-specific transcription was not detected (Fig. 4). Addition of partially purified Sp1 resulted in a significant activation of transcription. The pattern of Sp1-dependent transcripts was similar to that seen using the unfractionated extract, with a prominent start at the -55 site. As expected, Sp1 similarly stimulated the SV40 gene early promoter, but had no effect on the Ad-2 major late promoter. The *dhfr* template in this experiment was cleaved at -406, indicating that sequences upstream of this position are not required for Sp1-responsive transcription. Preliminary results, however, indicate the presence of additional Sp1 binding sequences in this upstream region. Moreover, in some experiments where sequences upstream to -900 were present, a higher basal level of transcription was seen in the absence of Sp1, reducing the net Sp1 stimulation. These results indicate that additional Sp1-dependent and independent transcription control elements are probably present in the region upstream of -406, and agree with earlier studies showing that this upstream region has promoter activity *in vivo* when fused to the DHFR structural gene⁴.

All mammalian *dhfr* genes so far examined contain potential Sp1 binding sites. The mouse⁴, human¹² and hamster (P. Mitchell and L. Chasin, personal communication) *dhfr* genes each have at least 5 GC box decanucleotide sequences within the 700 nucleotides upstream of the 5' end of the mRNA. Although interaction with Sp1 has been confirmed experimentally only for the mouse *dhfr* promoter, the apparent conservation of Sp1 sites provides additional evidence that the binding of this factor has a functional role.

Recent reports indicate the presence of potential Sp1 binding sites in the 5'-flanking regions of cellular genes in addition to

[‡] Present address: Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309, USA.

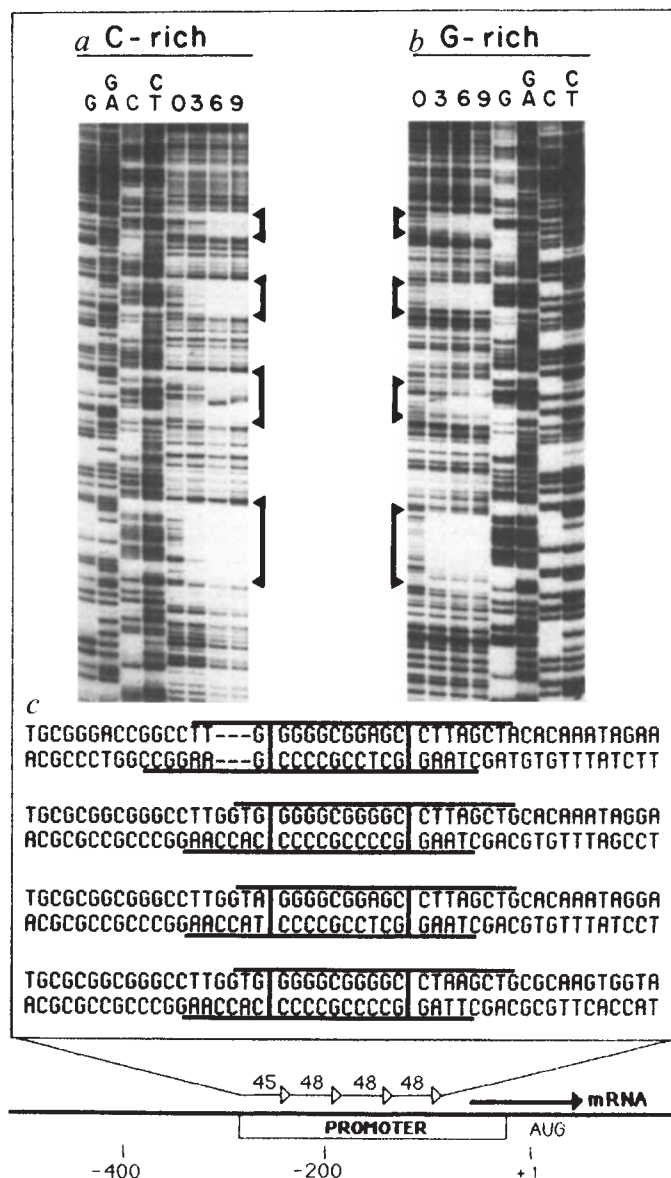


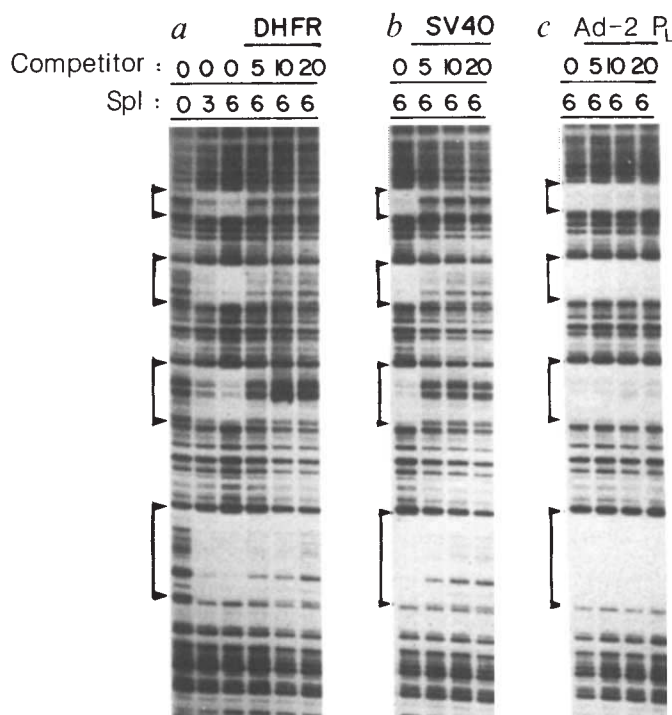
Fig. 1 Binding of Sp1 to a double-stranded DNA fragment containing the proximal portion of the *dhfr* promoter. **a**, 5' End of the C-rich strand labelled downstream of +6. DNA sequence markers were prepared by chemical cleavage²¹ (far left). Binding reactions contained 0, 3, 6 or 9 μ l of Sp1 as indicated above the lanes. Four discrete regions of protection are marked by vertical brackets on the right. **b**, Same fragment with 3' end of the G-rich strand labelled downstream of +6. DNA sequence markers and binding reactions are as shown on the left. The promoter region contains four 45–48-bp imperfect tandem repeats; the sequence is written at the bottom (**c**), aligned for maximum homology. Bars indicate areas protected from DNase and boxes enclose GC box decanucleotides. The consensus Sp1 binding decanucleotide derived from analysis of a number of viral and cellular promoters²² is:



The map indicates the position of the major mRNA initiation site relative to 45–48-bp repeats; the area marked as promoter is sufficient for *in vivo dhfr* expression. Numbering is relative to the *dhfr* initiation codon at +1.

Methods. DNA was labelled at a linker downstream of position +6, using polynucleotide kinase and [γ -³²P]ATP to label the 5' end of the C-rich strand, or reverse transcriptase and [α -³²P]-deoxynucleotide triphosphates to label the 3' end of the G-rich strand. DNA was cut with a second restriction enzyme in a linker upstream of -406, and the resulting fragment isolated. Pooled Sp1 fractions from the Sephacryl S-300 column¹ were used in DNase footprinting reactions^{2,9}. Binding incubation was at 0°C, with reactions warmed to room temperature immediately before DNase addition. Reactions contained 18 fmol of labelled probe and 1 μ g of native calf thymus DNA carrier.

Fig. 2 Binding of Sp1 to *dhfr* in the presence of competitor DNA. Three competitor DNA fragments were prepared: *dhfr*, a 412-bp fragment extending from -406 to +6; SV40, a 311-bp fragment (the EcoRII 'G' fragment) containing the functional early promoter and Sp1 binding sites; and Ad-2 P_L, a 290-bp fragment extending from 280 bp upstream of the adenovirus-2 major late initiation site to 10 bp downstream. The competitor fragment and native calf thymus DNA (0.5 μ g) were incubated with 0, 3 or 6 μ l of Sp1 for 10 min at 0°C, then 18 fmol of labelled *dhfr* fragment ('C-rich'; see Fig. 1) was added. The incubation was continued for a further 10 min at 0°C, then DNase was added and the products were analysed as described^{2,9}. The competitor fragment, when present, was in 5-, 10- or 20-fold molar excess over the labelled fragment, as indicated above each lane.



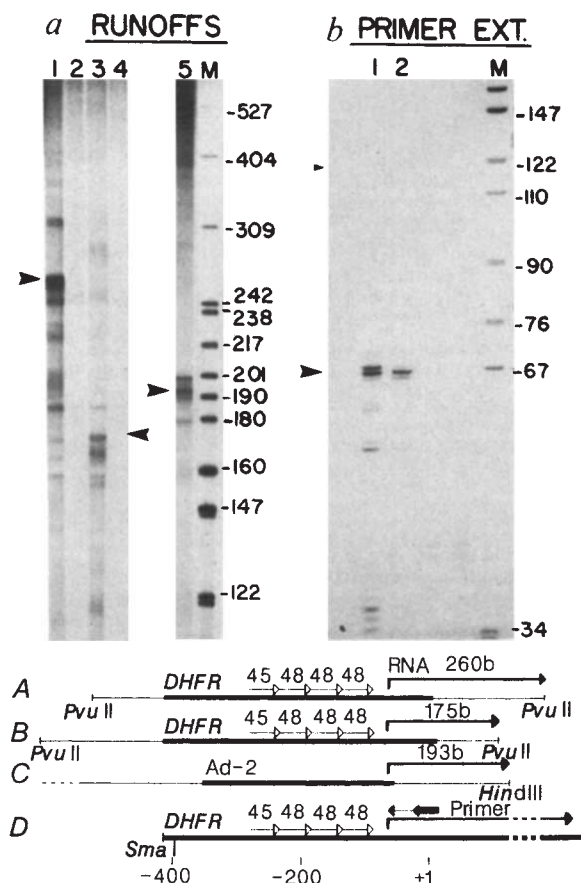


Fig. 3 Transcription of *dhfr*. *a*, *In vitro* runoff transcription. Lane 1, transcription directed by the *dhfr* promoter, template A in the diagram at the bottom. The arrow marks the major transcript migrating at 260 bases, relative to DNA size markers. Lane 2, as lane 1 with $1 \mu\text{g ml}^{-1}$ α -amanitin. Lane 3, transcription directed by the *dhfr* promoter, template B. The arrow marks the transcript at 175 bases. Lane 4, same with $1 \mu\text{g ml}^{-1}$ α -amanitin. Lane 5, transcription directed by adenovirus major late promoter, template C. The arrow marks the transcript at 195 bases. Lane M, pBR322/*Hpa*II size markers. *b*, Analysis by primer extension. Labeled 24-base synthetic primer has its 5' end at +11 in the *dhfr* coding region. Lane 1, *in vitro* transcription directed by the *dhfr* promoter, template D. The large arrow marks the major extension product, the small arrow the minor product. Lane 2, *in vivo* RNA from nuclei of *dhfr*-amplified cells. Lane M, pBR322/*Hpa*II size marker fragments. *c*, Diagram showing template DNAs. Thick lines, eukaryotic inserts; thin lines, attached vector sequences. **Methods.** Transcription reactions (50 μl) contained 15 μl of a nuclear extract prepared from human (HeLa) cells²³ and 500 ng of isolated *dhfr* promoter fragment (lanes 1–4) or 1,000 ng of plasmid containing the adenovirus major late promoter (lane 5). Reactions were run as described elsewhere¹. Runoff reactions contained 50 μCi [α -³²P] CTP and 5 μM unlabelled CTP, and were incubated for 60 min. Products were analysed on a thin 5% polyacrylamide-urea gel. Reactions for primer extension contained 250 μM nonradioactive CTP and were incubated for 30 min. *In vivo* RNA was isolated from nuclei of *dhfr*-amplified cell line 3T6R50 (ref. 24) as described elsewhere⁵. Hybridization and cDNA synthesis were as described previously², and products were analysed on a thin 8% polyacrylamide-urea gel.

dhfr. Several other genes involved in purine metabolism have sequences homologous to the consensus GC box decanucleotides, including those for mouse hypoxanthine phosphoribosyltransferase¹³, mouse adenine phosphoribosyltransferase¹⁴ and human adenosine deaminase¹⁵. Other genes with potential binding sites include rat type II procollagen¹⁶, several members of the human metallothionein family^{17,18}, human urokinase-plasminogen activator¹⁹ and hamster hydroxy-

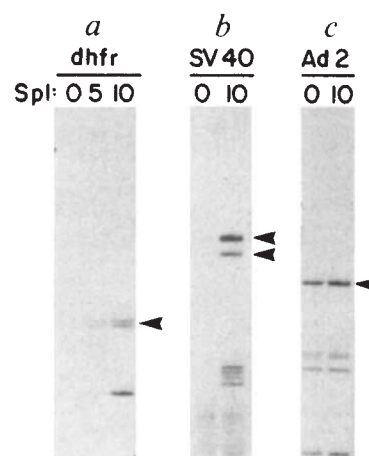


Fig. 4 Sp1-dependent *in vitro* transcription. Reactions contained 250 ng of promoter-containing linearized plasmid DNA, and 0, 5 or 10 μl of Sp1, as indicated. Separate complementary oligonucleotide primers were used for each template. Arrow in *a* marks the *dhfr* transcript initiating at about -55. Arrows in SV40 (*b*) and Ad-2 (*c*) mark transcripts initiating at the expected positions.

Methods. Reactions contained 9 μl of endogenous RNA polymerase II fraction and 3 μl of transcription factor Sp2, prepared from a whole-cell extract described previously¹. Sp1 was prepared by a new method (M. Briggs and R.T., unpublished) in which a nuclear extract was passed directly over an S-300 gel filtration column. Pooled fractions had approximately the same specific activity as those pooled from the S-300 column in the previous scheme. Similar stimulation of *dhfr* transcription has been seen with Sp1 fractions prepared by our previous method (not shown)¹.

methyl glutaryl CoA reductase²⁰. Because there is no known common pattern of regulation for this diverse group of genes, the possible role of Sp1 in their regulation is not immediately apparent. Perhaps the regulatory consequences of Sp1 binding depend on sequences adjacent to the Sp1 sites in a particular promoter, or are the result of interactions between Sp1 and other, as yet unidentified, factors. Further genetic analysis will help to elucidate the physiological role of Sp1 in the regulation of *dhfr* and other cellular genes.

We thank Pam Mitchell and Larry Chasin for providing the sequence of the 5'-flanking region of hamster *dhfr*. This research was funded by NIH grants to R.T. (CA25417) and R.T.S. (CA16318). W.S.D. was a Senior Postdoctoral Fellow of the American Cancer Society, California Division, Inc. S.S. was a predoctoral trainee supported by USPHS NIH grant GM07276.

Received 19 August; accepted 11 October 1985.

1. Dynan, W. S. & Tjian, R. *Cell* **32**, 669–680 (1983).
2. Dynan, W. S. & Tjian, R. *Cell* **35**, 79–87 (1983).
3. Dynan, W. S., Saffer, J., Lee, W. S. & Tjian, R. *T. Proc. natn. Acad. Sci. U.S.A.* (in the press).
4. McGrogan, M. M., Simonsen, C. C., Smouse, D. T., Farnham, P. J. & Schimke, R. T. *J. biol. Chem.* **260**, 2307–2314 (1985).
5. Farnham, P. J. & Schimke, R. T. *J. biol. Chem.* **260**, 7675–7680 (1985).
6. Tooze, J. *Molecular Biology of Tumor Viruses* 2nd edn, Pt 2, revised (Cold Spring Harbor Laboratory, New York, 1981).
7. Saffer, J. & Singer, M. *Nucleic Acids Res.* **12**, 4769–4788 (1984).
8. Farnham, P. J., Abrams, J. M. & Schimke, R. T. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3978–3982 (1985).
9. Gidoni, D., Dynan, W. S. & Tjian, R. *Nature* **312**, 409–413 (1984).
10. Jones, K. & Tjian, R. *Nature* **317**, 179–182 (1985).
11. Dynan, W. S. & Tjian, R. *Nature* **316**, 774–778 (1985).
12. Chen, M. J. *et al. J. biol. Chem.* **259**, 3933–3943 (1984).
13. Melton, D. W., Konecki, D. S., Brennand, J. & Caskey, C. T. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2147–2151 (1984).
14. Dush, M. K., Sikela, J. M., Khan, S. A., Tischfield, J. A. & Stambrook, P. J. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2731–2735 (1985).
15. Valerio, D. *et al. EMBO J.* **4**, 437–443 (1985).
16. Khono, K., Sullivan, M. & Yamada, Y. *J. biol. Chem.* **260**, 4441–4447 (1985).
17. Karin, M. & Richards, R. I. *Nature* **299**, 797–802 (1982).
18. Richards, R. I., Heguy, A. & Karin, M. *Cell* **37**, 263–272 (1984).
19. Riccio, A. *et al. Nucleic Acids Res.* **13**, 2759–2771 (1985).
20. Reynolds, A. B. *et al. Cell* **38**, 275–285 (1984).
21. Maxam, A. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).
22. Kadonaga, J., Jones, K. & Tjian, R. *Trends biochem. Sci.* (in the press).
23. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. *Nucleic Acids Res.* **11**, 1475–1489 (1983).
24. Brown, P. C., Beverley, S. M. & Schimke, R. T. *Molec. cell. Biol.* **1**, 1077–1083 (1981).
25. Sazer, S. & Schimke, R. T. *J. biol. Chem.* (in the press).